Segmentation and Representation for the Reuse of Skills Learned by Imitation

2012. 04. 18.

Intelligence and Control for Robots Laboratory,
Hanyang University, Korea
Il Hong Suh

Contents

- Definition of Skill Learning
- Four Stages of Skill Learning by Imitation
- “Big Five” Problems in Imitation
- State-of-the-Art in the Field of Skill Learning by Imitation
- Additionally Required Properties for Improving Reusability of Skills Learned
- Proposed Autonomous Segmentation Framework
  - Motivation
  - Conceptual Description
  - Quantitative Evaluation
- How can we reuse primitives well?
- Future Works

Definition of Skill Learning

Definition of "Skill"

- A special ability to something well, especially as gained by learning and practice
  < from dictionary of english language and culture, third edition >

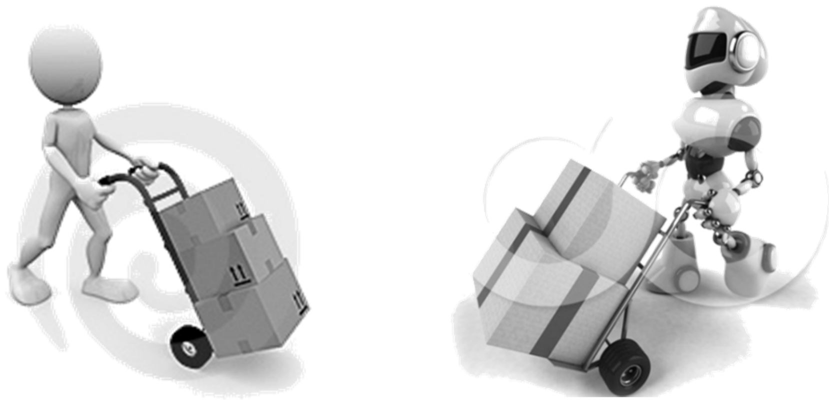
- A learned capacity to carry out pre-determined results often with the minimum outlay of time, energy, or both
  < from wikipedia>

- **In Robotics**
  - a sensory interactive robot control
    < J. S. Albus, "Mechanics of planning and problem solving the brain," Math. Bioscience, 1979>
  - appropriate goal-directed sequences of motor primitives
    < W. Erlhagen et. al., "Goal-directed imitation for robots: a bio-inspired approach to action understanding and skill learning," Robotics and Autonomous Systems, vol. 54, no. 5, pp.353-360, 2006>

Skill Learning

- Representing emergent behaviors (i.e. motor primitives)
- Representing sequences of the behaviors
- Refining the behaviors or their sequences by repeated practices and exercises

Skill Learning by Imitation



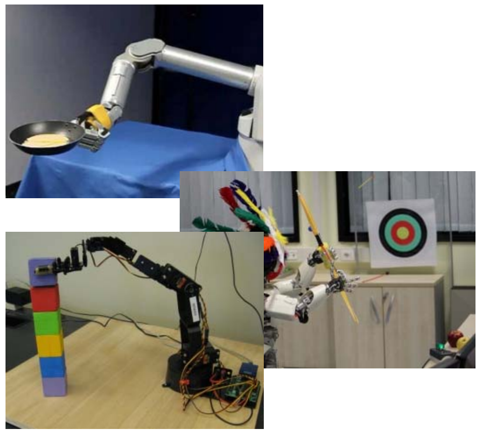
Imitation Learning

- **Learning behaviors that are stimulated by the perception of similar behaviors by another animal or person**
  <Albert Bandura; psychologist and philosopher (of action), 1925~>

- **A type of learning in which a naïve student copies an expert**
  - It can acquire novel skills by user-friendly interaction easily and quickly instead of programming new skills through machine commands.
  - It can promote to understand events of various types in the world easily.

Four Stages of Skill Learning by Imitation

1. Demonstration → 2. Imitation → 3. Reproduction → 4. Improvement



"Big Five": five central questions in imitation

I. Whom to Imitate


Who is good teacher?

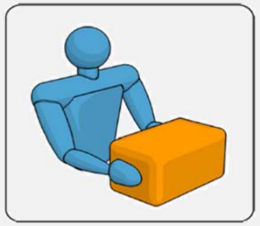
II. When to Imitate

The imitator has to decide on a suitable time to imitate.

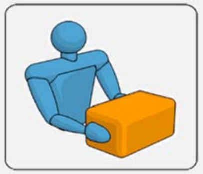
III. What to Imitate

States, Actions, Goals, Sequences?

IV. How to map observed to imitated behavior

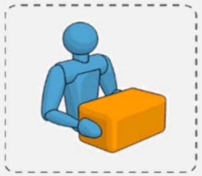


demonstration



reproduction in the same situation and the same embodiment

reproduction in the different situation



reproduction in the different embodiment

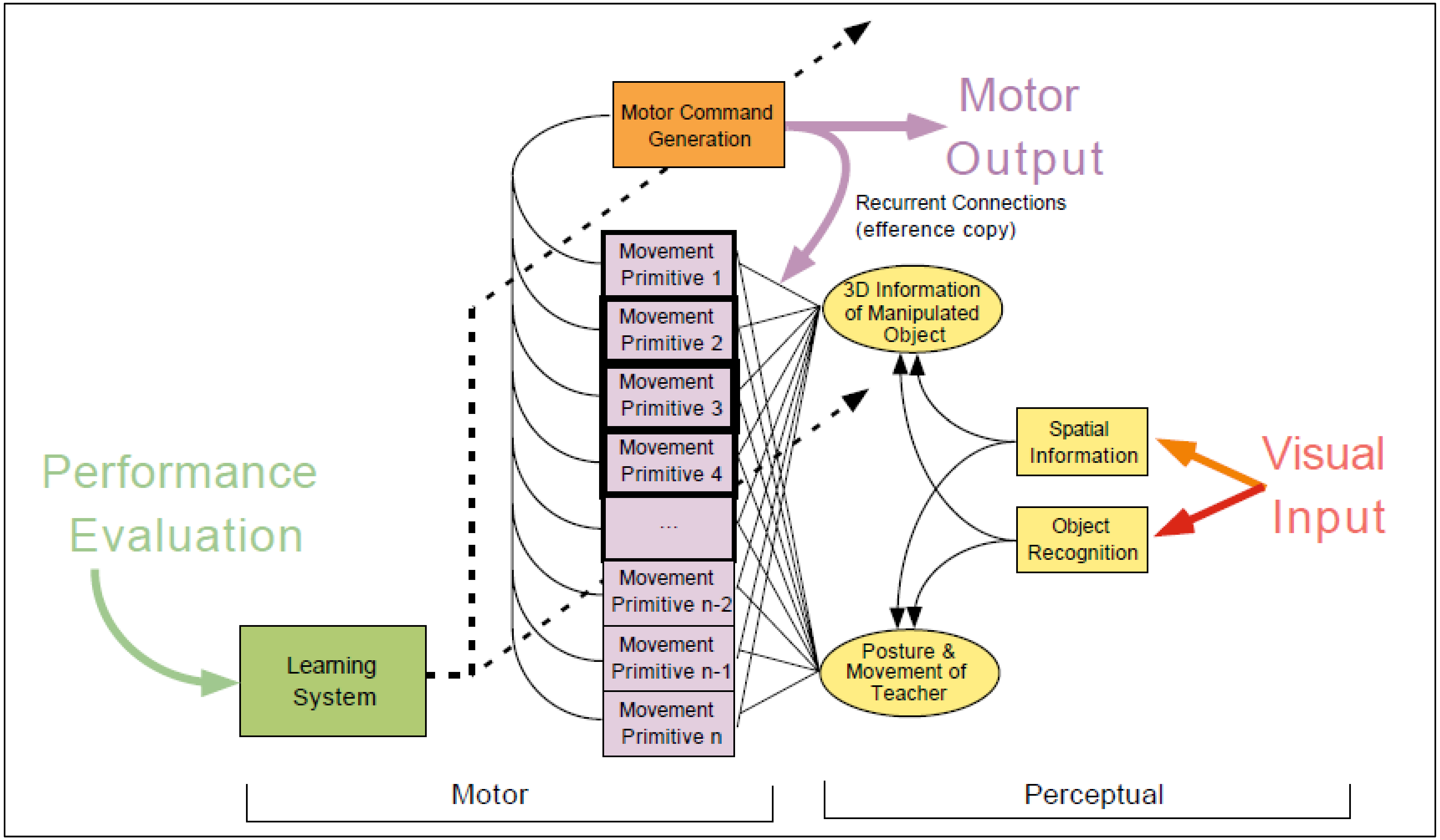
V. How to evaluate the success of imitation

Similarity

Success or failure by an external estimator

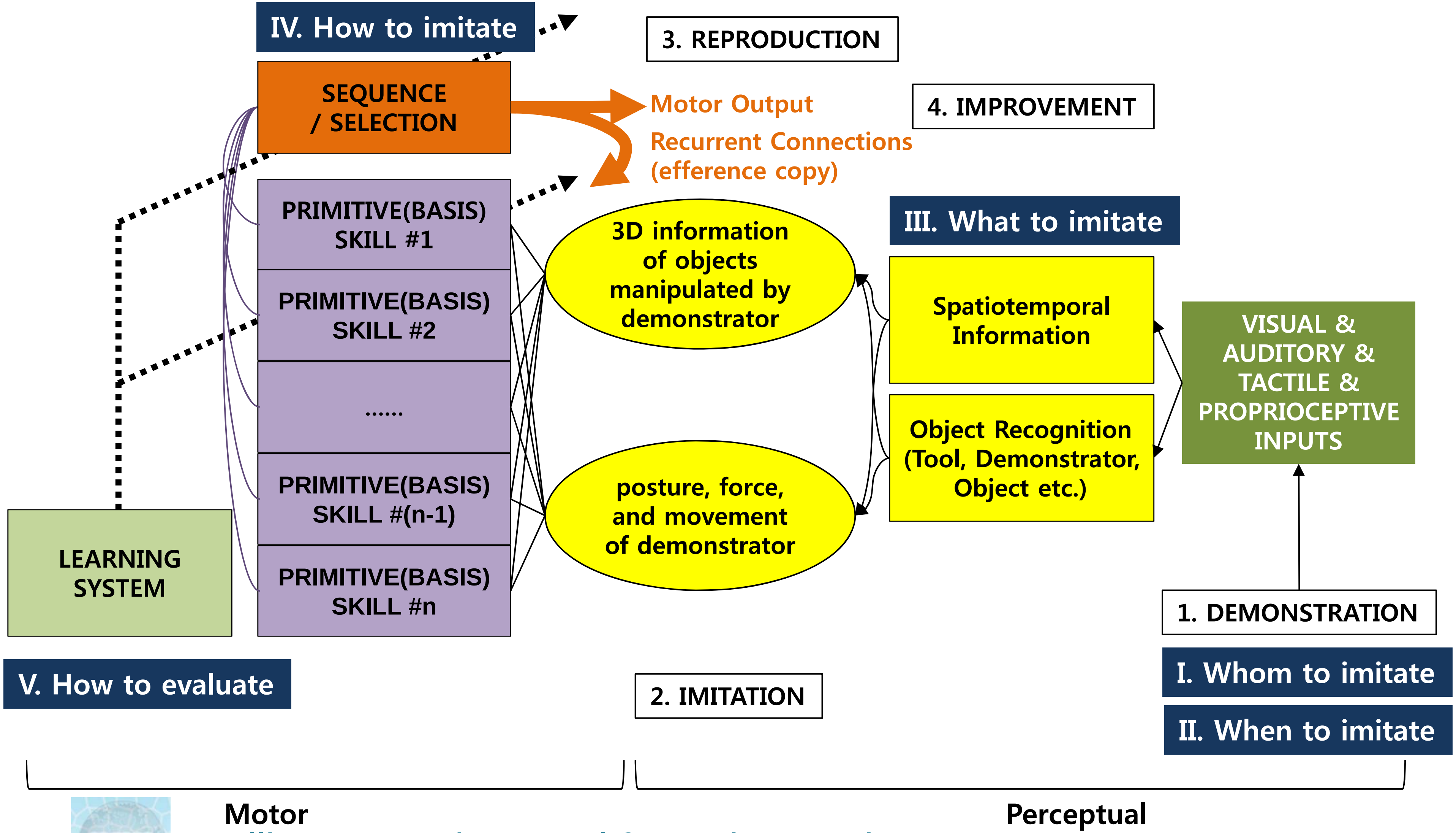
K. Dautenhahn and C. L. Nehaniv, "The Agent-Based Perspective on Imitation," Imitation in animals and artifacts, p`p.1-40, 2002, MIT Press

Conceptual Sketch on Skill Learning by Imitation

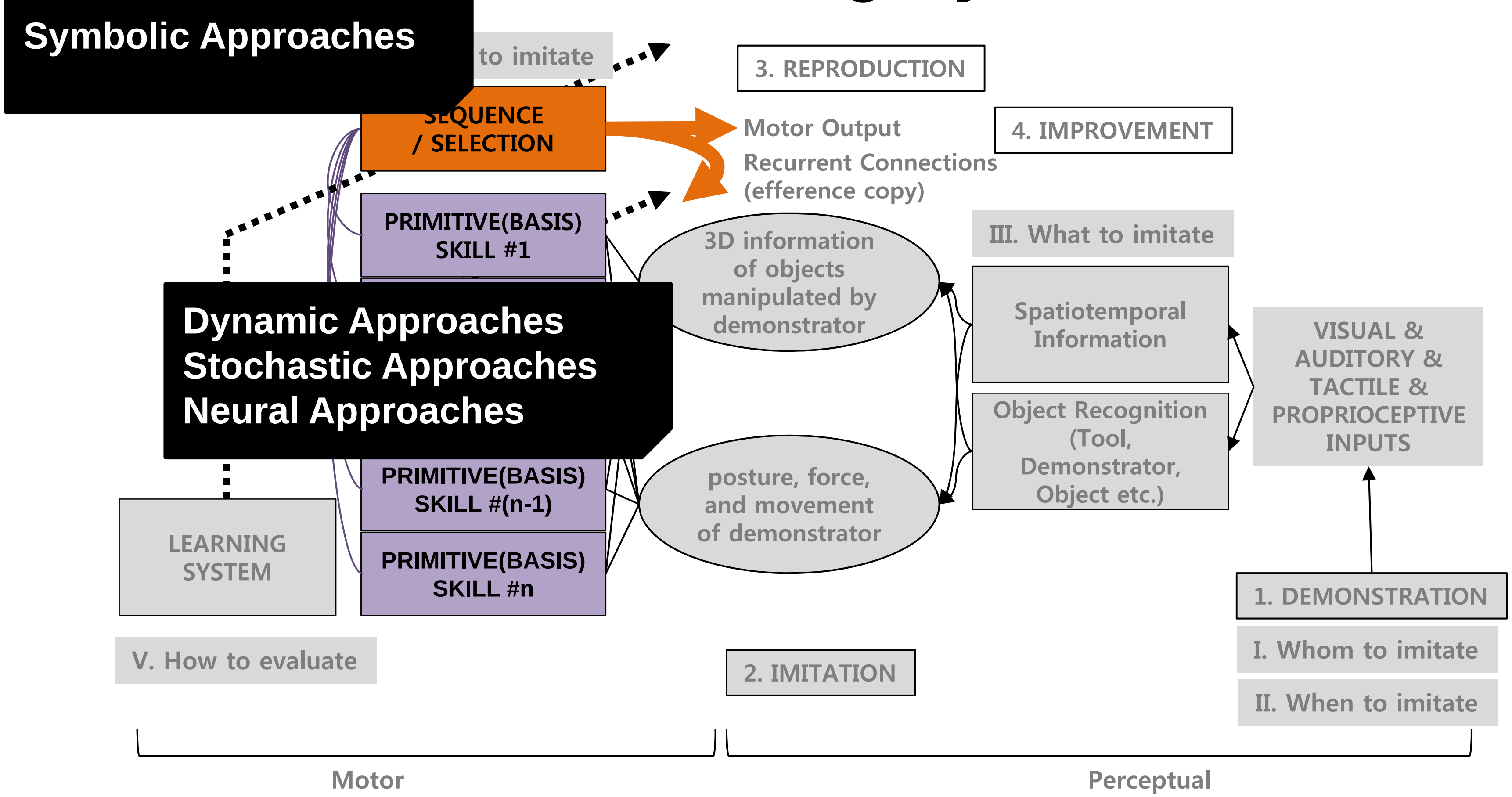


S. Schaal, "Is imitation learning the route to humanoid robots," Trends in cognitive sciences, vol. 3, no. 6, pp.233-242, 1999.

"Big Five" Problems Attached to Schaal's Conceptual Sketch



State-of-the-Art in the Field of Skill Learning by Imitation



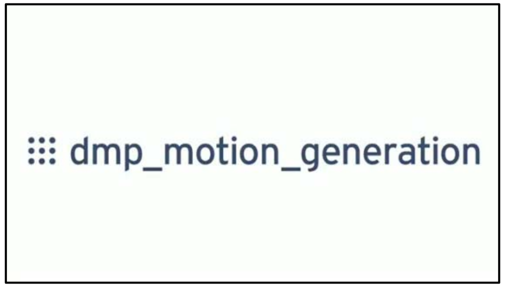
Symbolic Approaches : S. Ekvall (KTH), M. Pardowitz (Kalsruhe Univ.), J. Saunders (Hertfordshire Univ.)
Dynamic Approaches: A. Ijspeert (EPFL), S. Schaal (USC), C. G. Atkeson (GIT)
Stochastic Approaches: A. Billard (EPFL), D. H. Lee (TUM), S. Calinon (IIT)
Neural Approaches: E. Oztop (ATR), J. Ecety (Chicago Univ.), U. Demiris (South Kenshington)

State-of-the-Art:
Dynamic Approaches [1/2]

- Skill Learning Based on Dynamic Approach by Imitation

University of Southern California

[00:02:26]



[00:01:13]

Collaborative work

[00:00:44]

[00:02:05]

Max Planck Institute

[00:00:38]

[00:00:25]

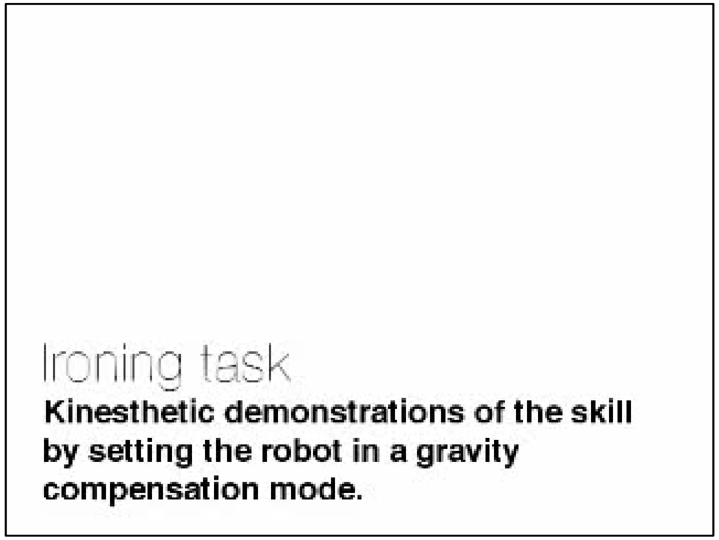
State-of-the-Art:
Dynamic Approaches [2/2]

- Skill Learning Based on Dynamic Approach by Imitation

Willow Garage

Italian Institute of Technology

[00:01:51]



[00:02:28]

State-of-the-Art:
Stochastic Approaches [1/2]

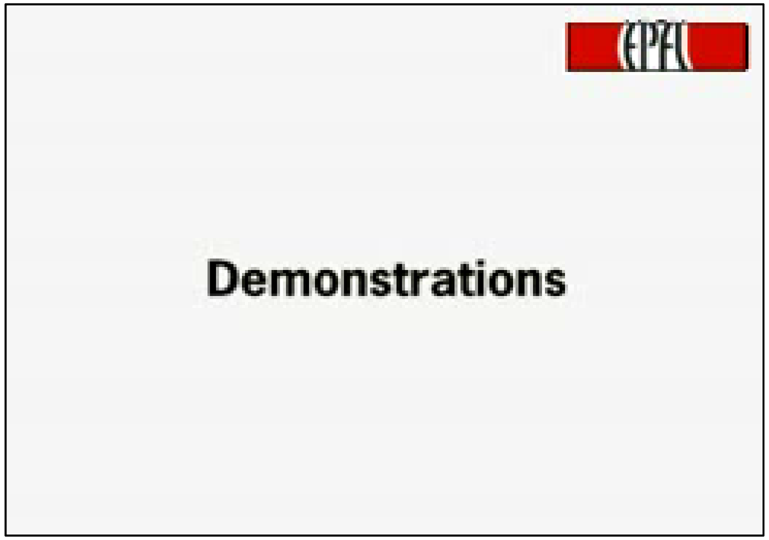
- Skill Learning Based on Stochastic Approach by Imitation

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Technische Universität München

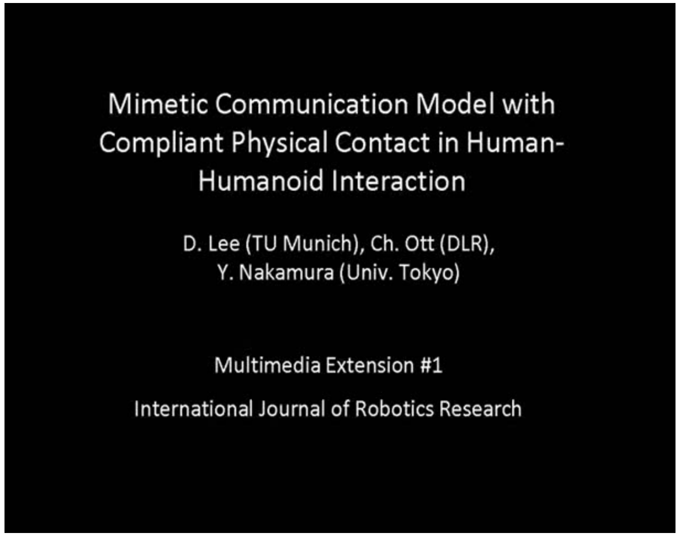
École polytechnique fédérale de Lausanne

- Based on GMM/GMR -



[00:02:29]

- Based on HMM -



[00:02:40]

State-of-the-Art:
Stochastic Approaches [2/2]

- Skill Learning Based on Stochastic Approach by Imitation

Karlsruhe Institute of Technology

- Based on HMM -

[00:00:55]

Italian Institute of Technology

- Based on HSMM, GMM/GMR -

[00:01:44]

[00:01:27]

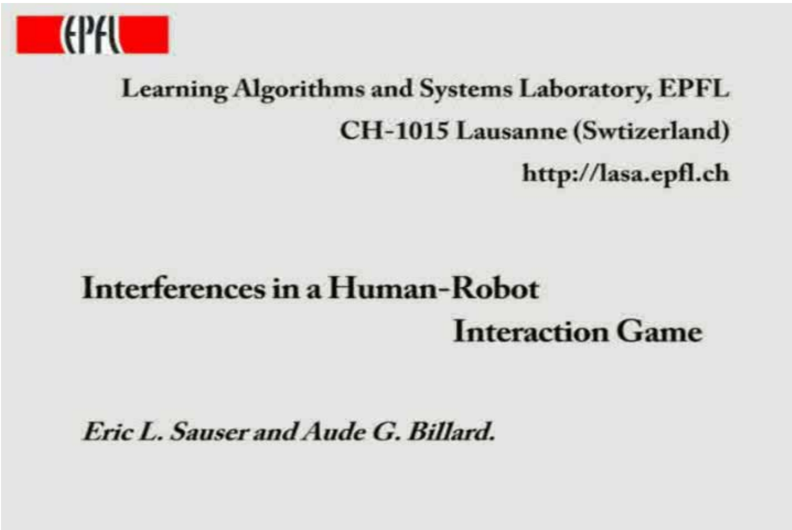
State-of-the-Art: Neural Approaches [1/1]

- Skill Learning Based on Neural approach by imitation

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

École polytechnique fédérale de Lausanne

Advanced Telecommunications Research Institute International 株式会社 国際電気通信基礎技術研究所



[00:02:59]

[00:01:56]

State-of-the-Art:
Skill Improvement [1/1]

- Skill Improvement by Reinforcement Learning

Max Planck Institute

[00:00:15]

[00:00:07]

[00:00:40]

Italian Institute of Technology

[00:01:38]

[00:02:07]

Willow Garage

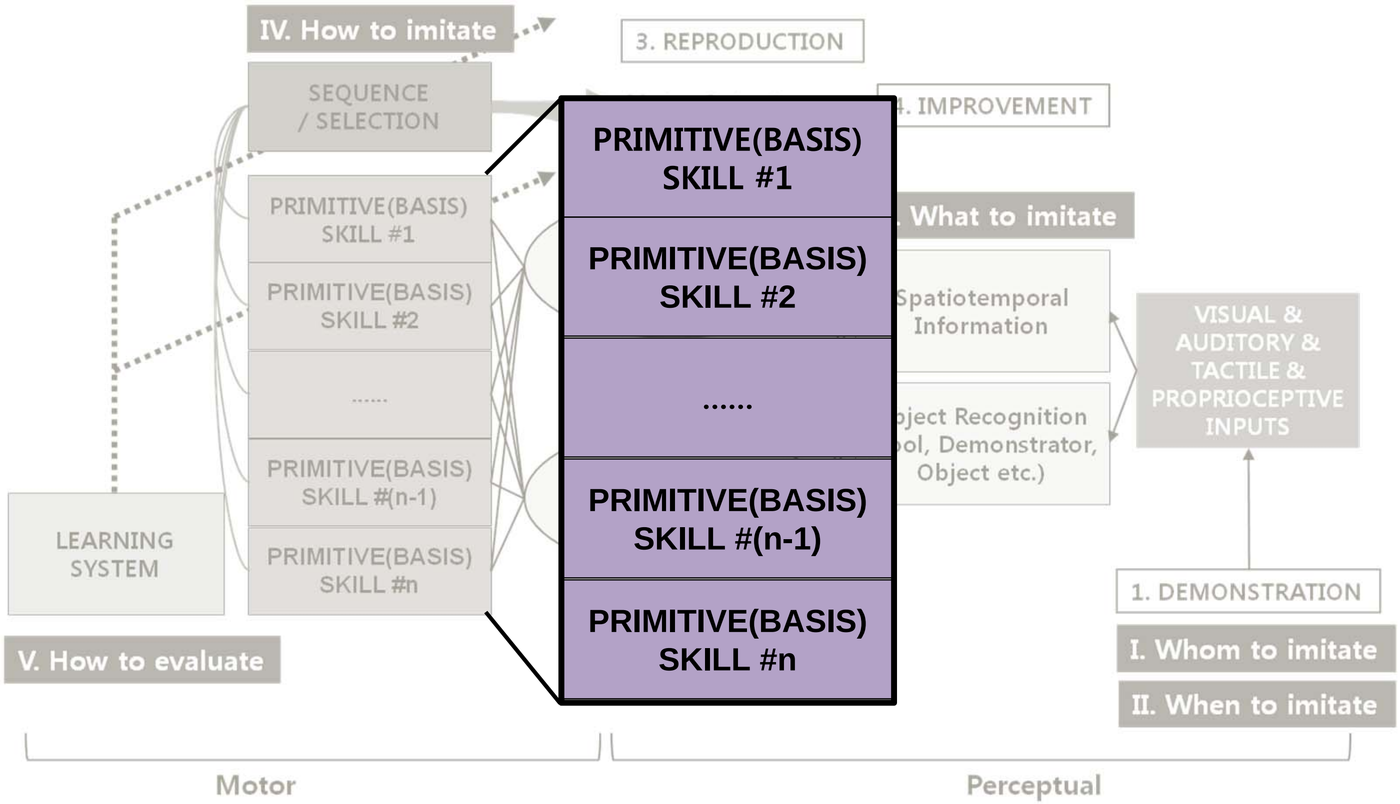
[00:01:42]

State-of-the-Art: Summary

- State-of-the-Art in the field of Skill Learning by Imitation

Skill Learning by Imitation	
Approaches	1. Symbolic Approaches 2. Dynamical Approaches 3. Stochastic Approaches 4. Neural Approaches
Properties	1. Easy programming 2. Ability to generalize to new situations 3. Ability against perturbations 4. Skill Improvement by self-demonstration
Additionally Required Properties	Improvement of Reusability

Additionally Required Properties for Improving Reusability of Skills Learned



Additionally Required Properties for Reuse of Skills Learned by Imitation	- **Autonomous Segmentation for Learning Primitives** - **Reorganization of Primitive Skills for Alternative Solutions** - **Classification of Primitives** - **Generalization of Primitives**

State-of-the-Art of Segmentation Approaches

Researcher	Affiliation	Methods
Supervised Approaches : how can we predefine the primitives?		
Unsupervised Approaches : how can we tune the values?		

Motivation of Autonomous Segmentation Framework

- Reorganization of New Sentences using Words

sentences "HEISABOY" "SHEISNOTAGIRL"

words

HE	IS	A	BOY

SHE	IS	NOT	A	GIRL

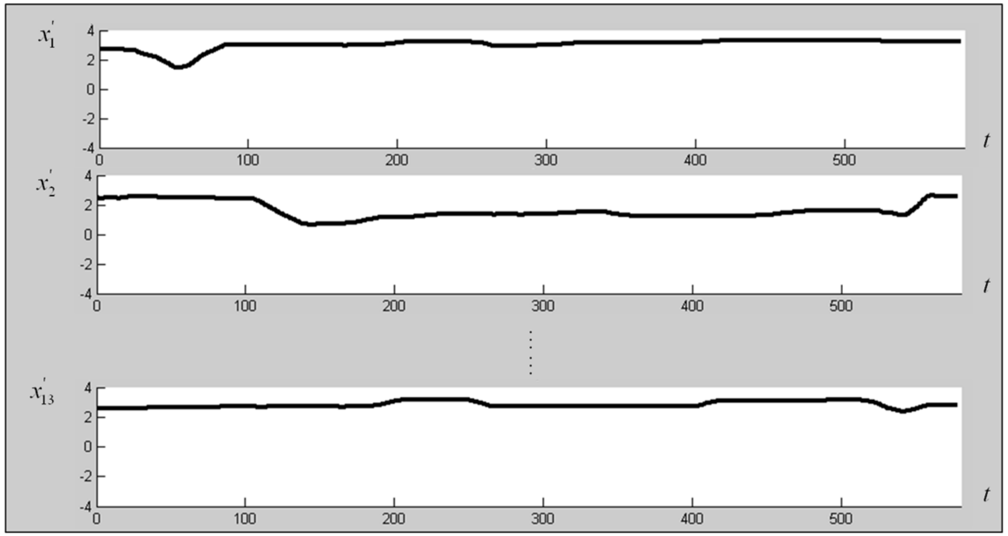
reorganization

New sentence "HE IS NOT A GIRL"

Autonomous Segmentation Framework
: Conceptual Description [1/5]

- How many primitives are contained in this continuous trajectories?

< Joint Trajectories Extracted from a Humanoid Robot >



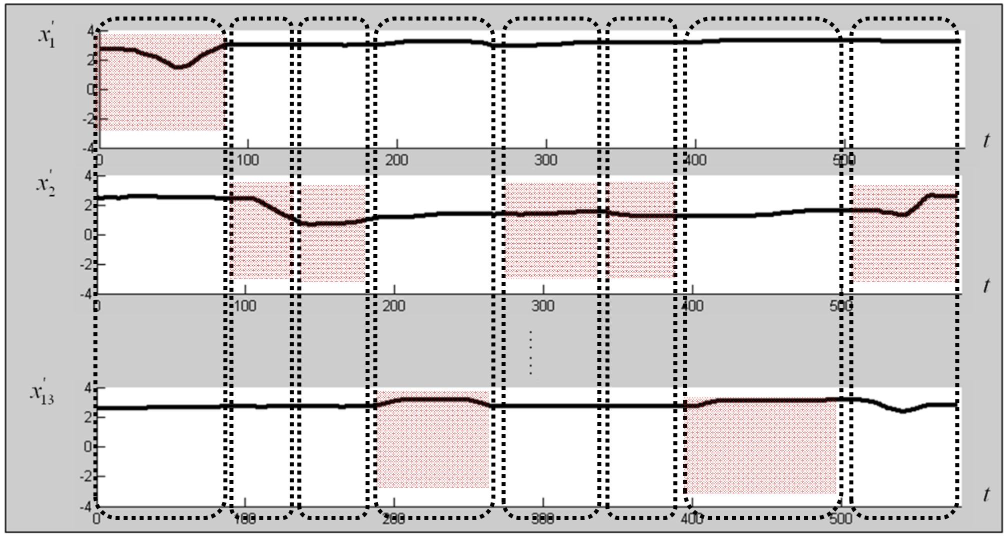
changing local movement?
(e.g., velocities, directions, dynamics, relations etc.)

<S. H. Lee, I. H. Suh, S. Calinon, and R. Johansson, "Autonomous Segmentation Framework for Alternative Solutions in Manipulation Task," submitted to an international journal, 2012 >

Autonomous Segmentation Framework : Conceptual Description [2/5]

- How many primitives are contained in this continuous trajectories?

< Joint Trajectories Extracted from a Humanoid Robot >

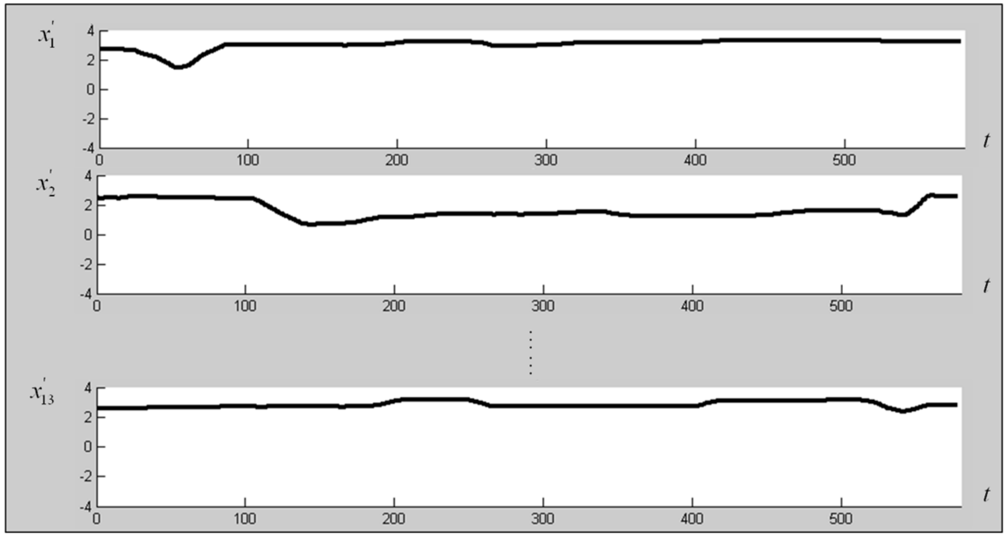


If a human intuitively divides this continuous trajectories according to changing local directions of the trajectories…

Autonomous Segmentation Framework : Conceptual Description [3/5]

- How many primitives are contained in this continuous trajectories?

< Joint Trajectories Extracted from a Humanoid Robot >



Gaussian Mixture Model (GMM)

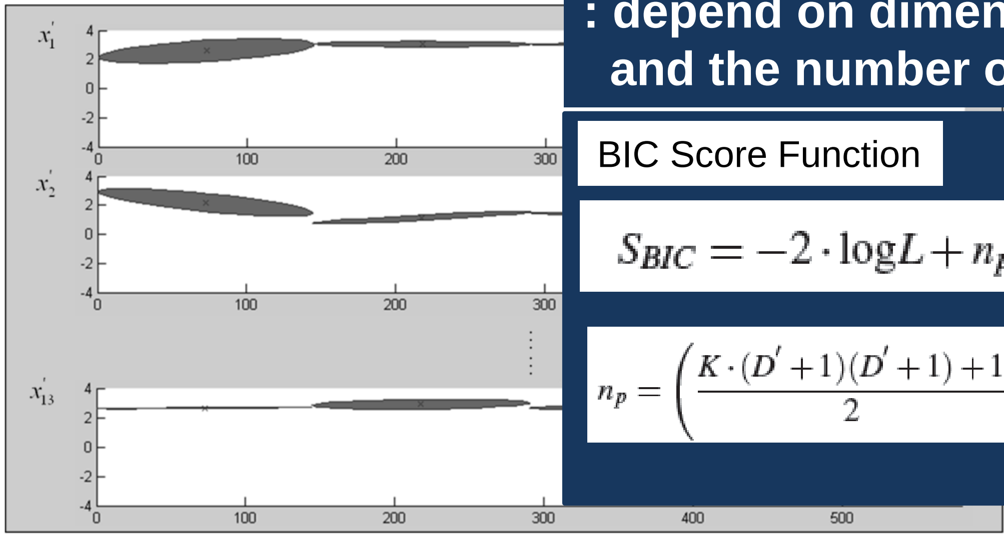
- Representing continuous trajectories as a GMM provides a way of encoding **the local directions and the local relations** (i.e. correlation and variances) among the variables taking part in the traj

a change of the local directions and relations in the GMM domain → a segmentation point

Autonomous Segmentation Framework
: Conceptual Description [4/5]

- Then, how can the number of Gaussians be determined in the GMM?

→ ***Bayesian information criterion (BIC) following r***



the estimated GMM by using the number of Gaussians automatically determined by BIC

The Score of "BIC"
: depend on dimension of variables and the number of Gaussians

BIC Score Function

$$S_{BIC} = -2 \cdot \log L + n_p \cdot \log(N),$$

$$n_p = \left(\frac{K \cdot (D' + 1)(D' + 1) + 1}{2}\right) + (K-1) + (K \cdot (D' + 1)),$$

the dimensionalities of variables	↑	and	↓
the number of Gaussians	↓		↑

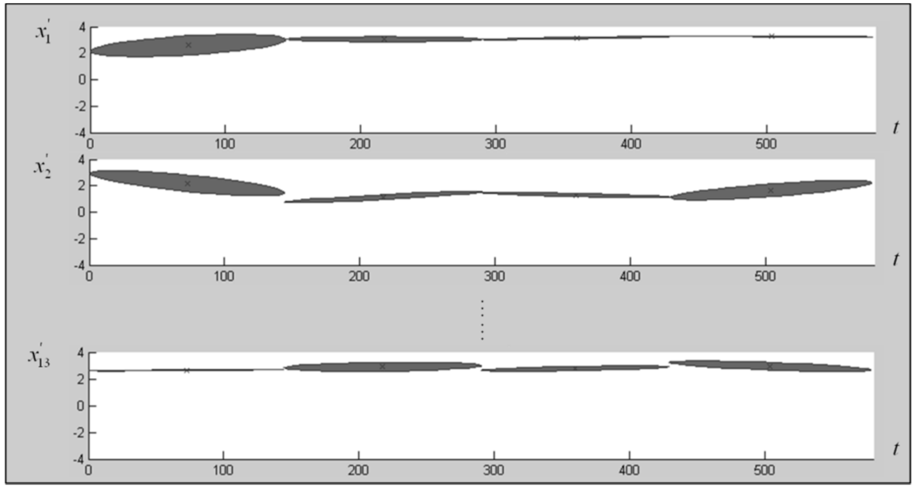
Strategy of this framework

: find as ***many meaningful primitives*** as possible by reducing the dimensionalities of variables

Principal Component Analysis (PCA)

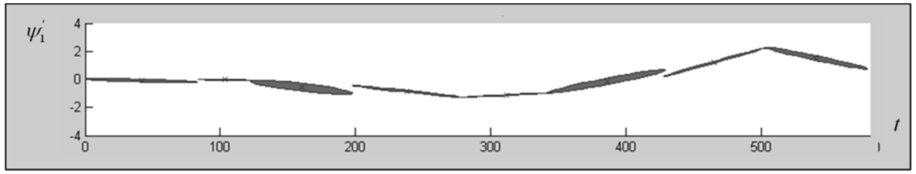
Autonomous Segmentation Framework
: Conceptual Description [5/5]

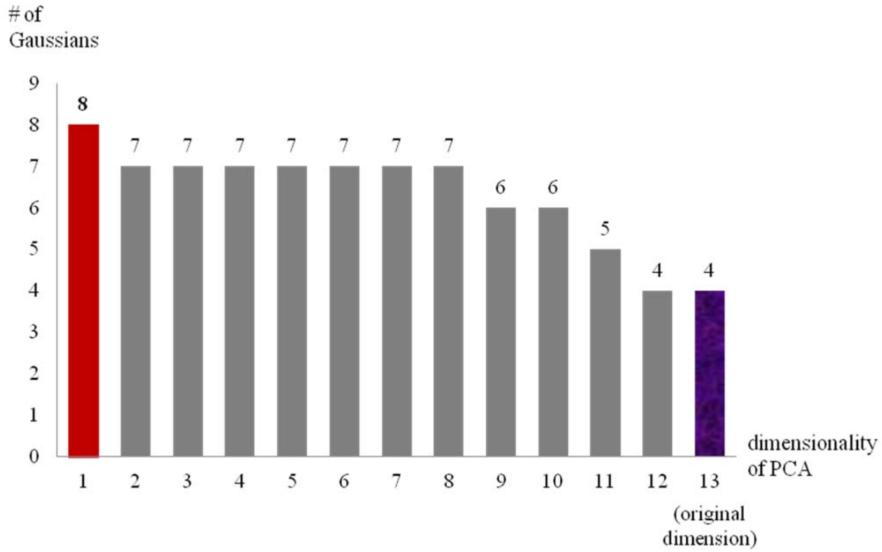
< Joint Trajectories Extracted from a Humanoid Robot >



< in original space >

[motion trajectories in the dimensional space reduced by PCA]

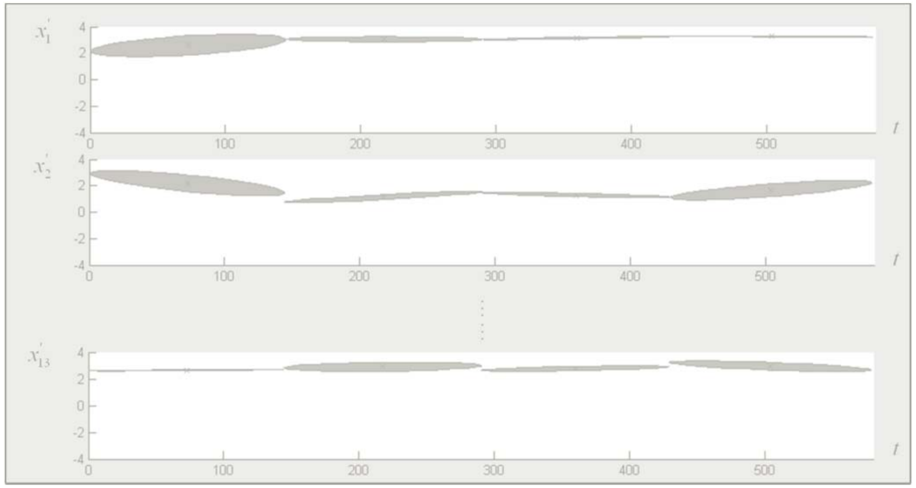




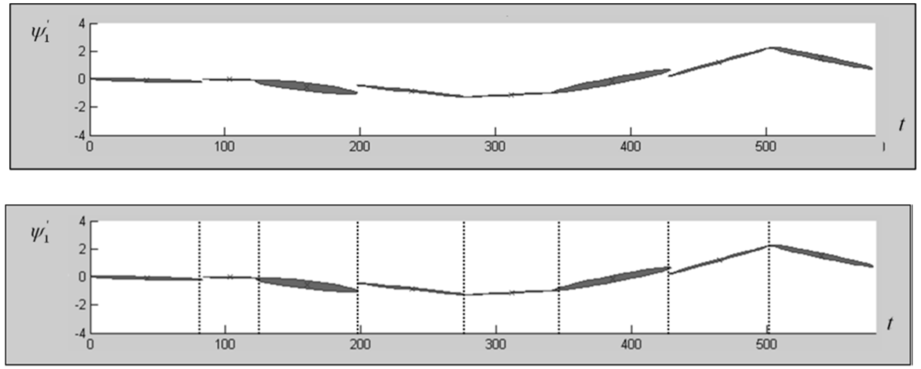
The number of Gaussians estimated according to the dimensionality of PCA when using BIC

Autonomous Segmentation Framework
: Conceptual Description [5/5]

< Joint Trajectories Extracted from a Humanoid Robot >



[motion trajectories in the dimensional space reduced by PCA]

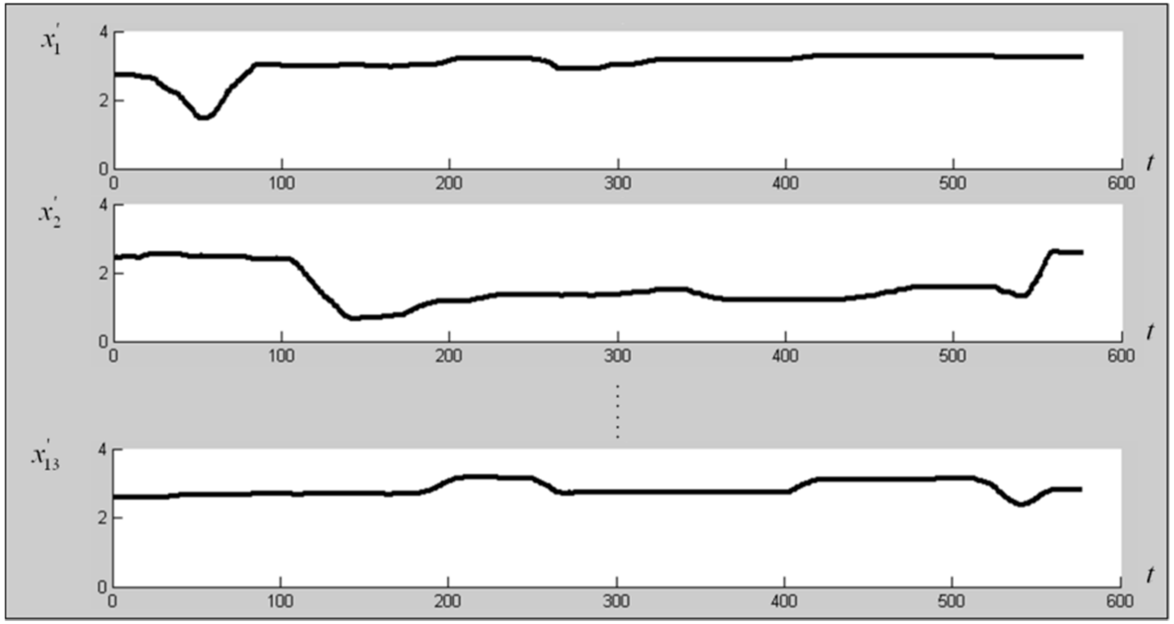


Changes of the local directions and relations in the GMM domain (the set of segmentation point)

Temporally overlapping points In-between two consecutive Gaussians

Autonomous Segmentation Framework : “Cooking Rice” Task [1/9]

< Joint trajectories extracted from a single demonstration in the task of cooking rice >



Cooking Rice

[PROCEDURE]

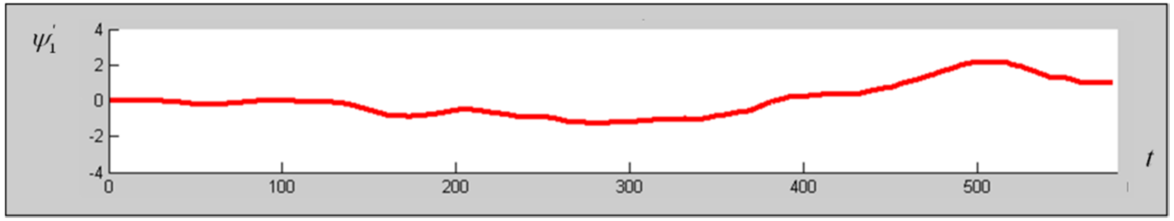
1. The robot **lifts** the pot, which is attached to the right hand toward kitchen board.
2. The robot **scoops** some grains of rice from a rice bowl using a spoon attached to its left hand.
3. The rice is **delivered** from the bowl to the pot.
4. The robot **pours** the rice into the pot.
5. The robot **stirs** the rice in the pot using the spoon.
6. The pot is **put on** the stove.

<Kinesthetic Teaching Process>

[00:00:12]

Autonomous Segmentation Framework : "Cooking Rice" Task [2/9]

< Motion trajectories in the dimensional space reduced by PCA >



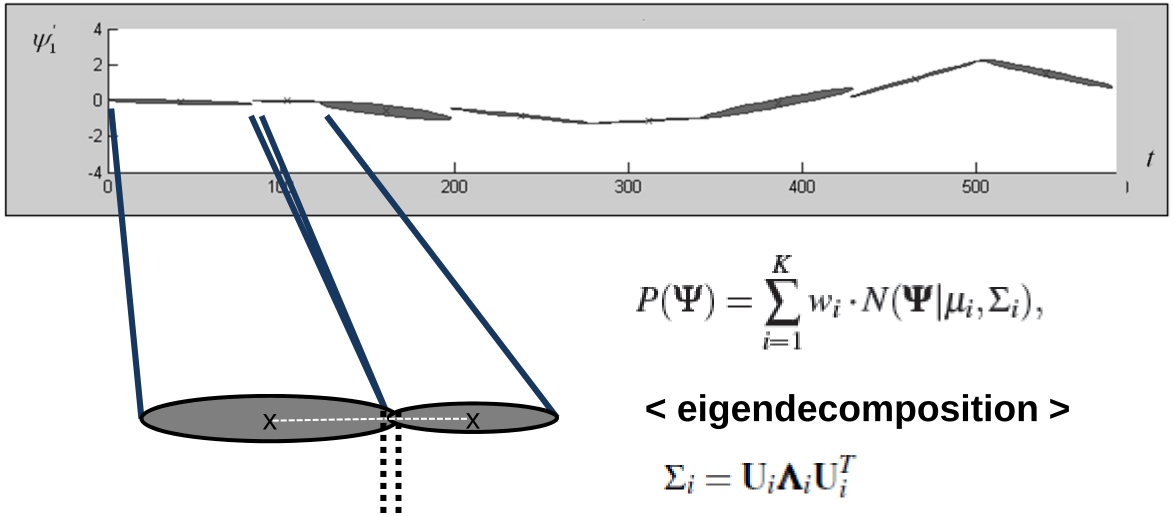
Cooking Rice

[PROCEDURE]

1. The robot **lifts** the pot, which is attached to the right hand toward kitchen board.
2. The robot **scoops** some grains of rice from a rice bowl using a spoon attached to its left hand.
3. The rice is **delivered** from the bowl to the pot.
4. The robot **pours** the rice into the pot.
5. The robot **stirs** the rice in the pot using the spoon.
6. The pot is **put on** the stove.

Autonomous Segmentation Framework : “Cooking Rice” Task [3/9]

< GMM that consists of eight Gaussians estimated by BIC and EM >



temporally
overlapping region

$$P(\boldsymbol{\Psi}) = \sum_{i=1}^{K} w_i \cdot N(\boldsymbol{\Psi}|\mu_i,\Sigma_i),$$

< eigendecomposition >

$$\Sigma_i = \mathbf{U}_i \boldsymbol{\Lambda}_i \mathbf{U}_i^T$$

- Geometrically, the i^th Gaussian $N(\boldsymbol{\Psi}|\mu_i,\Sigma_i)$ is identified with the distribution in which the normal distribution $N(\mathbf{I},0)$ is scaled by $\boldsymbol{\Lambda}_i^{1/2}$, rotated by $\mathbf{U}_i$, and translated by μ_i.
- The geometrical sizes of eigenvectors on the Guassian are therefore calculated using square root of the eigenvalue $\boldsymbol{\Lambda}_i$.

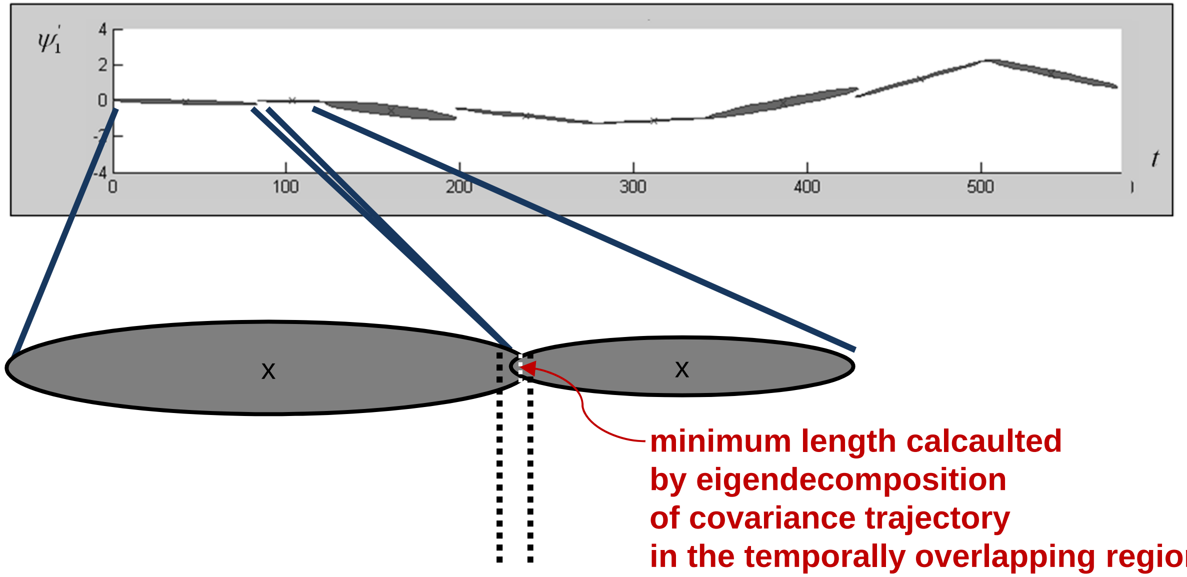
Cooking Rice

[PROCEDURE]

1. The robot **lifts** the pot, which is attached to the right hand toward kitchen board.
2. The robot **scoops** some grains of rice from a rice bowl using a spoon attached to its left hand.
3. The rice is **delivered** from the bowl to the pot.
4. The robot **pours** the rice into the pot.
5. The robot **stirs** the rice in the pot using the spoon.
6. The pot is **put on** the stove.

Autonomous Segmentation Framework : “Cooking Rice” Task [4/9]

< GMM that consists of eight Gaussians estimated by BIC and EM >



temporally
overlapping region

< Gaussian Mixture Regression>

$$\Sigma_{\Psi'}(t) = \sum_{i=1}^{K} h_i^2(t)\left(\Sigma_{i,\Psi'} - \Sigma_{i,\Psi' t}\Sigma_{i,t}^{-1}\Sigma_{i,t\Psi'}\right),$$: covariance trajectory

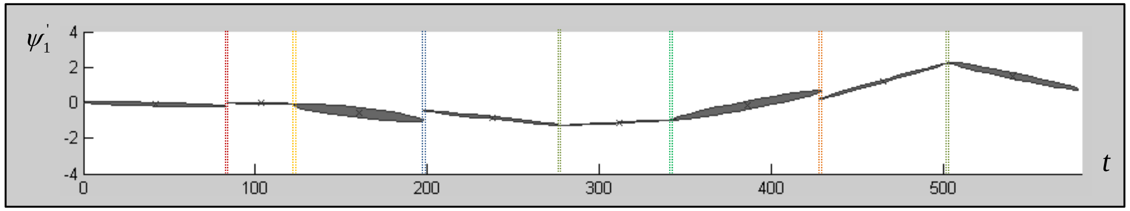
Cooking Rice

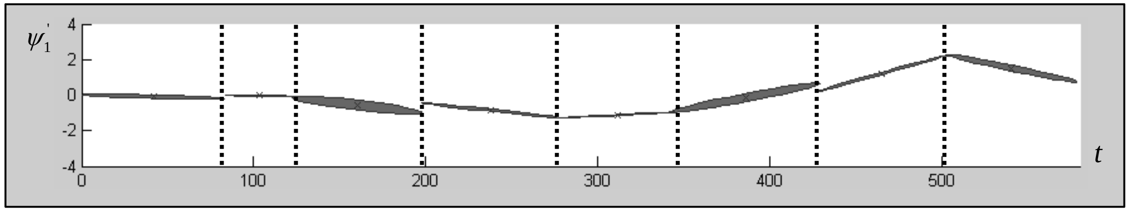
[PROCEDURE]

1. The robot **lifts** the pot, which is attached to the right hand toward kitchen board.
2. The robot **scoops** some grains of rice from a rice bowl using a spoon attached to its left hand.
3. The rice is **delivered** from the bowl to the pot.
4. The robot **pours** the rice into the pot.
5. The robot **stirs** the rice in the pot using the spoon.
6. The pot is **put on** the stove.

Autonomous Segmentation Framework : "Cooking Rice" Task [5/9]

< Temporally overlapping regions estimated by geometrical interpretation of the Gaussians >





< Segmentation points estimated by weights along the time component of the GMM >

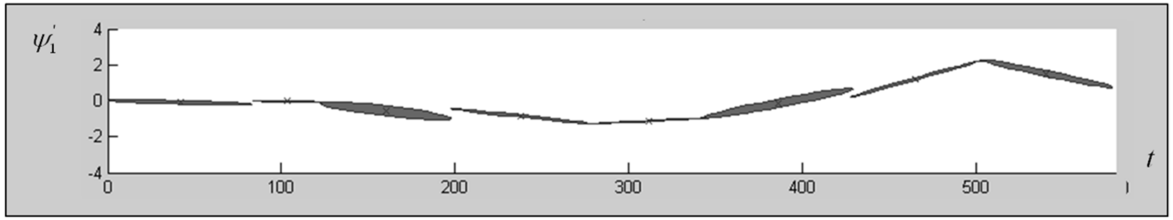
Cooking Rice

[PROCEDURE]

1. The robot **lifts** the pot, which is attached to the right hand toward kitchen board.
2. The robot **scoops** some grains of rice from a rice bowl using a spoon attached to its left hand.
3. The rice is **delivered** from the bowl to the pot.
4. The robot **pours** the rice into the pot.
5. The robot **stirs** the rice in the pot using the spoon.
6. The pot is **put on** the stove.

Autonomous Segmentation Framework : “Cooking Rice” Task [6/9]

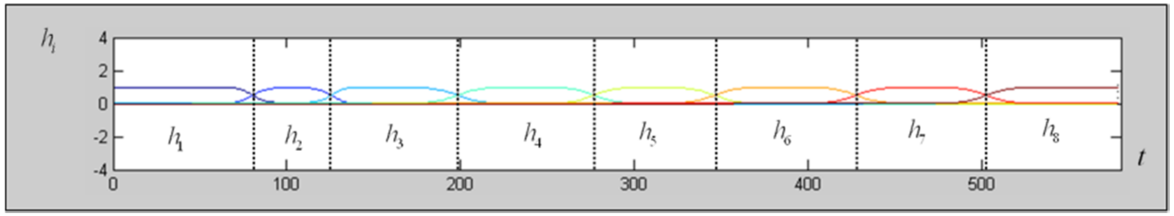
< GMM that consists of eight Gaussians estimated by BIC and EM >



$$P(\Psi)=\sum_{i=1}^{K} w_i \cdot N(\Psi|\mu_i,\Sigma_i),$$

Other method

< Weights estimated along the time component of the GMM and intersections by the weights >



$$h_i(t)=\frac{w_i N(t;\mu_{i,t},\Sigma_{i,t})}{\sum_{k=1}^{K} w_k N(t;\mu_{k,t},\Sigma_{k,t})},$$

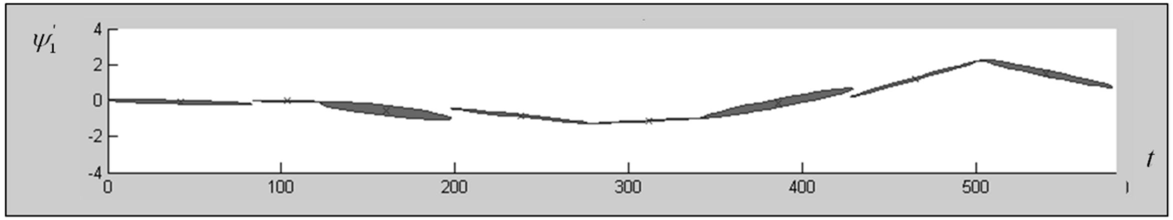
Cooking Rice

[PROCEDURE]

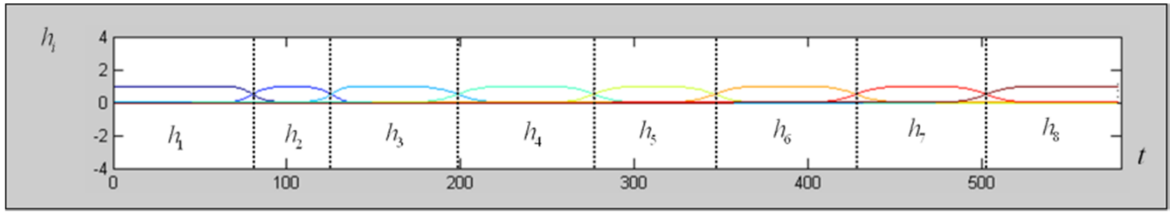
1. The robot **lifts** the pot, which is attached to the right hand toward kitchen board.
2. The robot **scoops** some grains of rice from a rice bowl using a spoon attached to its left hand.
3. The rice is **delivered** from the bowl to the pot.
4. The robot **pours** the rice into the pot.
5. The robot **stirs** the rice in the pot using the spoon.
6. The pot is **put on** the stove.

Autonomous Segmentation Framework : “Cooking Rice” Task [7/9]

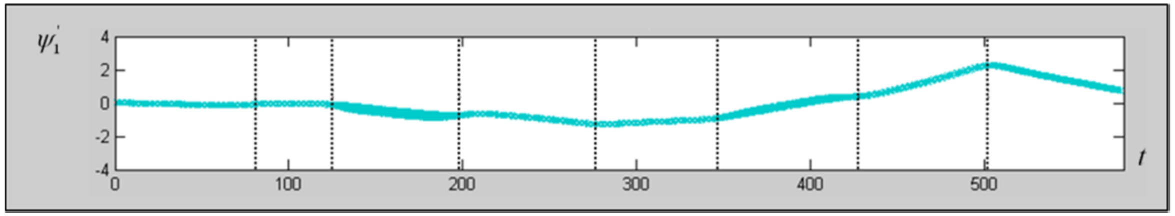
< GMM that consists of eight Gaussians estimated by BIC and EM >



< Weights estimated along the time component of the GMM and intersections by the weights >



< Continuous trajectories generalized by GMR process temporally overlapping points in between consecutive Gaussians >



Cooking Rice

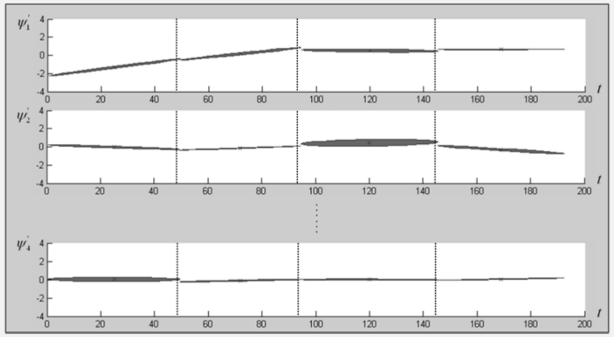
[PROCEDURE]

1. The robot **lifts** the pot, which is attached to the right hand toward kitchen board.
2. The robot **scoops** some grains of rice from a rice bowl using a spoon attached to its left hand.
3. The rice is **delivered** from the bowl to the pot.
4. The robot **pours** the rice into the pot.
5. The robot **stirs** the rice in the pot using the spoon.
6. The pot is **put on** the stove.

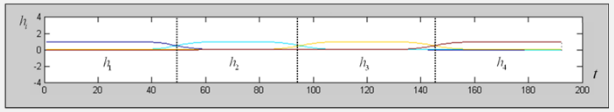
Autonomous Segmentation Framework : “Cutting a Food Item” Task [8/9]

Segmentation Point Detection / Reorganization / GMR

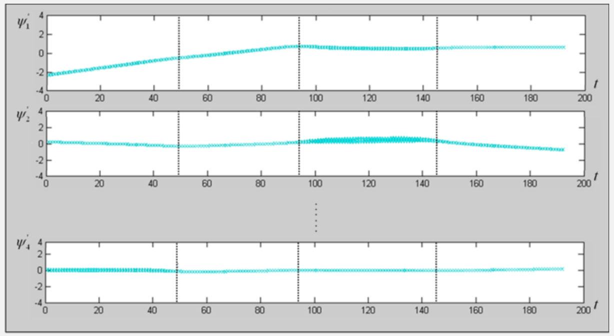
< GMM that consists of four Gaussians and temporally overlapping points in-between consecutive Gaussians >



< Weights estimated along the time component of the GMM and intersections by the weights >



< Continuous trajectories generalized by GMR process when sequentially organizing eight Gaussians >



Cutting a Food Item

[PROCEDURE]

1. The robot cuts a food item on a cutting board once only using a knife attached to its left hand.
2. The robot pushes the cut items into the pot attached to its right hand.

<Kinesthetic Teaching Process>

[00:00:15]

Segmentation Results Acquired by Autonomous Segmentation Framework [9/9]

- Two Cooking Tasks : 1. cooking rice and 2. cutting a food item

Segmentation Results

[Task of Cooking Rice]

[00:00:19]

Eights Segments : [LiftingPot], [LiftingSpoon], [ApproachingRiceBowl], [ScoopingRice], [DeliveringRice], [PouringRice], [StirringRice], and [PuttingOnStove].

[Task of Cutting a Food Item]

[00:00:08]

Four Segments : [LiftingKnife], [CuttingFoodItem], [PositioningForPushing], [PushingFoodItem]

Quantitative Evaluation of Autonomous Segmentation Framework [1/5]

< Four episodes opened from TUM Kitchen dataset >

episode [ID0-0]

[00:01:06]

episode [ID0-2]

[00:00:54]

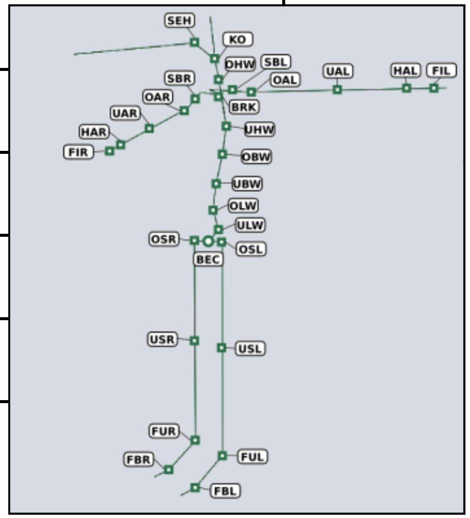
episode [ID0-11]

[00:01:35]

episode [ID0-12]

[00:01:03]

Labels of Nine Primitives Segmented by A. Yao	
Left Arm & Right Arm	**Trunk**
CarryingWhileLocomoting (CWL)	**StandingStill (Standing)**
Reaching	**HumanWalkingProcess (Walking)**
TakingSomething (Taking)	
OpeningADoor (Opening)	
LoweringAnObject (Lowering)	
ClosingADoor (Closing)	
ReleasingGraspofSomething (Releasing)	



28 body parts x 3 (x, y, z)
= 84-dimensional motion capture data
recorded at 25Hz

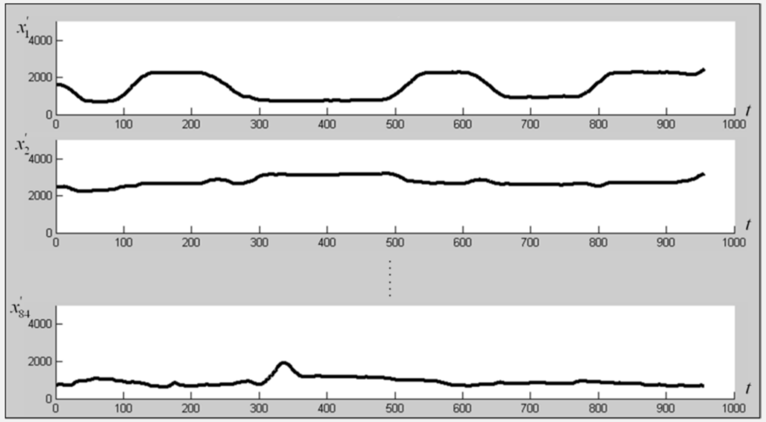
Quantitative Evaluation of Autonomous Segmentation Framework [2/5]

Labels of Nine Primitives Segmented by A. Yao		Labels of Sixteen Basis Primitives Segmented by Our Method	
Left Arm & Right Arm	**Trunk**	**Left Arm & Right Arm**	**Trunk**
CarryingWhileLocomoting (CWL)	StandingStill (Standing)	Meaningless Movment (M_M)	Standing
Reaching	HumanWalkingProcess (Walking)	StretchingToGrasp (G_Stretching)	WalkingForward (F_Walking)
TakingSomething (Taking)		GraspingObjects (Grasping)	WalkingBackward (B_Walking)
OpeningADoor (Opening)		StretchingToOpenDoor (O_Stretching)	WalkingSideways (S_Walking)
		FoldingToOpenDoor (O_Folding)	
LoweringAnObject (Lowering)		StretchingToRelease (R_Stretching)	TurningUsingLeftFoot (L_Turning)
ClosingADoor (Closing)		StretchingToCloseDoor (C_Stretching)	TurningUsingRightFoot (R_Turning)
		FoldingToCloseDoor (C_Folding)	
ReleasingGraspofSomething (Releasing)		ReleasingObjects (Releasing)	

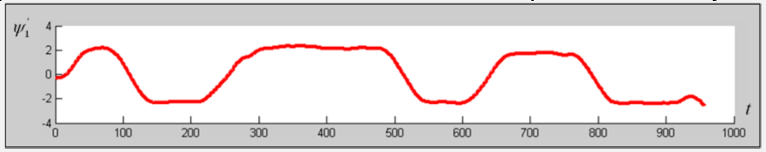
Quantitative Evaluation of Autonomous Segmentation Framework [3/5]

Autonomous Segmentation Process using TUM episode [ID0-2]

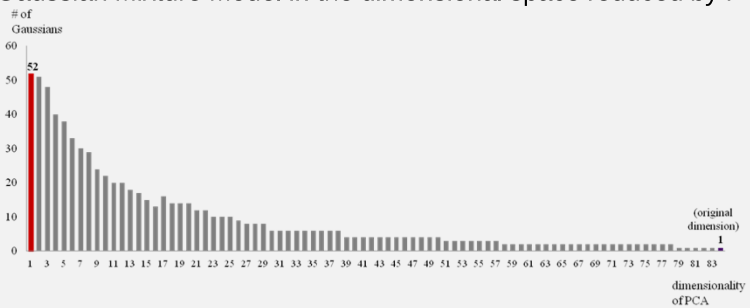
[Gaussian Mixture Model in the dimensional space reduced by PCA]



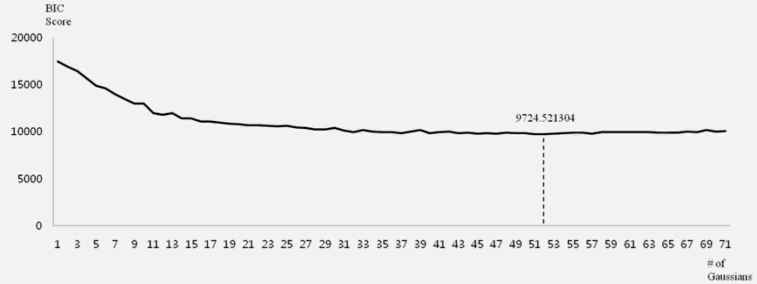
[Gaussian Mixture Model in the dimensional space reduced by PCA]



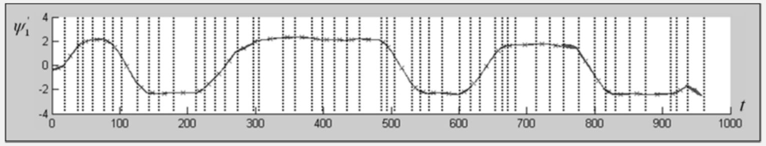
[Gaussian Mixture Model in the dimensional space reduced by PCA]



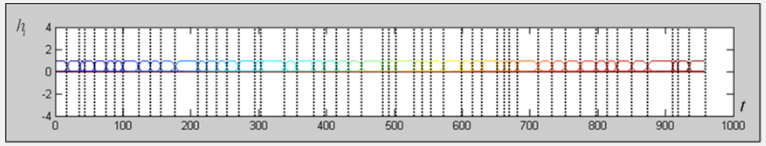
[Gaussian Mixture Model in the dimensional space reduced by PCA]



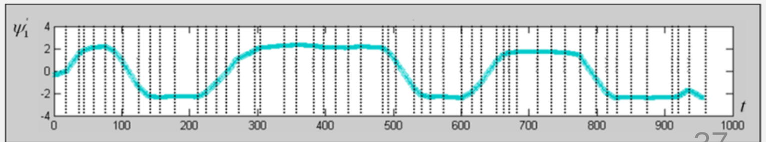
[Gaussian Mixture Model in the dimensional space reduced by PCA]



[Gaussian Mixture Model in the dimensional space reduced by PCA]



[Gaussian Mixture Model in the dimensional space reduced by PCA]



Quantitative Evaluation of Autonomous Segmentation Framework [4/5]

No.	A. Yao's method					Our proposed method				
	Left Hand	Right Hand	Trunk	Time	# of segments	Left Hand	Right Hand	Trunk	Time	# of segments
1	CWL	CWL	STANDING	001 ~ 010	01	M_M	M_M	STANDING	001 ~ 015	01
2	CWL REACHING	CWL REACHING	WALKING WALKING	011 ~ 028 029 ~ 042	02 03	G_STRETCHING	M_M	F_WARKING	016 ~ 040	02
3	REACHING REACHING	TAKING TAKING	W ST					TANDING TANDING	041 ~ 057 058 ~ 075	03 04
4	TAKING CWL	TAKING CWL	ST ST					_TURNING	076 ~ 087	05
5	CWL	CWL	W					_TURNING _WALKING	088 ~ 098 099 ~ 123	06 07
6	CWL	LOWERING	WALKING	126 ~ 136	09	M_M	R_STRETCHING	F_WALKING	124 ~ 142	08
7	CWL CWL	LOWERING RELEASING	STANDING STANDING	137 ~ 167 168 ~ 175	10 11	M_M M_M	R_STRETCHING RELEASING	STANDING STANDING	143 ~ 156 157 ~ 177	09 10
8	CWL	LOWERING	STANDING	176 ~ 203	12	M_M	R_STRETCHING	STANDING	178 ~ 210	11
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
35	LOWERING LOWERING RELEASING CWL CWL	CWL LOWERING LOWERING LOWERING RELEASING	STANDING STANDING STANDING STANDING STANDING	826 ~ 828 829 ~ 833 834 ~ 889 890 ~ 899 900 ~ 919	50 51 52 53 54	R_STRETCHING R_STRETCHING RELEASING M_M	M_M R_STRETCHING R_STRETCHING RELEASING	STANDING STANDING STANDING STANDING	828 ~ 850 815 ~ 873 874 ~ 910 911 ~ 918	47 48 49 40
36	CWL	RELEASING	WALKING	920 ~ 931	55	M_M	RELEASING	B _WALKING	919 ~ 934	51
37	CWL	CWL	WALKING	932 ~ 957	56	M_M	M_M	F_WALKING	935 ~ 957	52

when timing differences between the starting and ending points in the segments are allowed to 0 to 10 frames (i.e. 0.0~0.4 sec)

Quantitative Evaluation of Autonomous ...mentation Framework [5/5]

No			method			Our proposed method				
			Trunk	Time	# of segments	Left Hand	Right Hand	Trunk	Time	# of segmetns
1			NDING	001 ~ 010	01	M_M	M_M	STANDING	001 ~ 015	01
2	**CWL** **REACHING**	**CWL** **REACHING**	**WALKING** **WALKING**	**011 ~ 028** **029 ~ 042**	**02** **03**	**G_STRETCHING**	**M_M**	**F_WARKING**	**016 ~ 040**	**02**
3	REACHING REACHIN	TAKING	WALKING	043 ~ 071	04 05	G_STRETCHING GRASPING	G_STRETCHING	STANDING	041 ~ 057 058 ~ 075	03 04
4	TAKING CWL			~ 083	06 07	GRASPING			076 ~ 087	05
5	CWL			~ 125	08	M_M M_M			088 ~ 098 099 ~ 123	06 07
6	CWL			~ 136	09	M_M			124 ~ 142	08
7	CWL CWL			~ 167 ~ 175	10 11	M_M M_M			143 ~ 156 157 ~ 177	09 10
8	CWL	LOWERING	STANDING	176 ~ 203	12	M_M	R_STRETCHING	STANDING	178 ~ 210	11
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
35	LOWERING LOWERING RELEASING CWL CWL	CWL LOWERING LOWERING LOWERING RELEASING	STANDING STANDING STANDING STANDING STANDING	826 ~ 828 829 ~ 833 834 ~ 889 890 ~ 899 900 ~ 919	50 51 52 53 54	R_STRETCHING R_STRETCHING RELEASING M_M	M_M R_STRETCHING R_STRETCHING RELEASING	STANDING STANDING STANDING STANDING	828 ~ 850 815 ~ 873 874 ~ 910 911 ~ 918	47 48 49 40
36	CWL	RELEASING	WALKING	920 ~ 931	55	M_M	RELEASING	B _WALKING	919 ~ 934	51
37	CWL	CWL	WALKING	932 ~ 957	56	M_M	M_M	F_WALKING	935 ~ 957	53

Quantitative Evaluation of Autonomous Segmentation Framework [5/5]

No.	A. Yao's method					Our proposed method				
	Left Ha			Time	# of segments	Left Hand			Time	# of segments
1	CWL			1 ~ 010	01	M_M			1 ~ 015	01
2	CWL REACHI			1 ~ 028 29 ~ 042	02 03	G_STRETCHING			6 ~ 040	02
3	REACHI REACHI			3 4 ~ 071	04 05	G_STRETCHING GRASPING			1 ~ 057 8 ~ 075	03 04
4	TAKING CWL	TAKING CWL	STANDING STANDING	072 ~ 083 084	06 07	GRASPING	GRASPING	L_TURNING	076 ~ 087	05
5	**CWL**	**CWL**	**WALKING**	**085 ~ 125**	**08**	**M_M** **M_M**	**M_M** **M_M**	**R_TURNING** **F_WALKING**	**088 ~ 098** **099 ~ 123**	**06** **07**
6	CWL	LOWERING	WALKING	126 ~ 136	09	M_M	R_STRETCHING	F_WALKING	124 ~ [illegible]	08
7	CWL CWL	LOWERING RELEASING	STANDING STANDING	137 ~ 167 168 ~ 175	10 11	M_M M_M			3 ~ 156 7 ~ 177	09 10
8	CWL	LOWERING	STANDING	176 ~ 203	12	M_M			8 ~ 210	11
⋮	⋮	⋮	⋮	⋮	⋮	⋮			⋮	⋮
35	LOWERING LOWERING RELEASING CWL CWL	CWL LOWERING LOWERING LOWERING RELEASING	STANDING STANDING STANDING STANDING STANDING	826 ~ 828 829 ~ 833 834 ~ 889 890 ~ 899 900 ~ 919	50 51 52 53 54	R_STRETCHING R_STRETCHING RELEASING M_M	RELEASING	STANDING	8 ~ 850 5 ~ 873 4 ~ 910 911 ~ 918	47 48 49 40
36	CWL	RELEASING	WALKING	920 ~ 931	55	M_M	RELEASING	B _WALKING	919 ~ 934	51
37	CWL	CWL	WALKING	932 ~ 957	56	M_M	M_M	F_WALKING	935 ~ 957	[illegible]

Quantitative Evaluation of Autonomous Segmentation Framework [5/5]

No.	A. Yao's method					Our proposed method				
	Left Hand	Right Hand	Trunk	Time	# of segments	Left Hand	Right Hand	Trunk	Time	# of segments
1	CWL	CWL	STANDING	001 ~ 010	01	M_M	M_M	STANDING	001 ~ 015	01
2	CWL REACHING	CWL REACHING	WALKING WALKING	011 ~ 028 029 ~ 042	02 03	G_STRETCHING	M_M	F_WARKING	016 ~ 040	02
3	REACHING REACHING	TAKING TAKING	WALKING STANDING	043 044 ~ 071	04 05	G_STRETCHING GRASPING	G_STRETCHING GRASPING	STANDING STANDING	041 ~ 057 058 ~ 075	03 04
4	TAKING CWL	TAKING CWL	STANDING STANDING	072 ~ 083 084	06 07	GRASPING	GRASPING	L_TURNING	076 ~ 087	05
5	CWL	CWL	WALKING	085 ~ 125	08	M_M M_M	M_M M_M	R_TURNING F_WALKING	088 ~ 098 099 ~ 123	06 07
6	CWL	LOWERING	WALKING	126 ~ 136	09	M_M	R_STRETCHING	F_WALKING	124 ~ 142	08
7	CWL CWL			~ 167 ~ 175	10 11	M_M M_M			3 ~ 156 7 ~ 177	09 10
8	CWL			~ 203	12	M_M			8 ~ 210	11
⋮	⋮			⋮	⋮	⋮			⋮	⋮
35	LOWERING LOWERING RELEASING CWL CWL	 LOWERING RELEASING	 STANDING STANDING	~ 828 ~ 833 ~ 889 890 ~ 899 900 ~ 910	50 51 52 53 54	R_STRETCHING R_STRETCHING RELEASING M_M	 RELEASING	 STANDING	8 ~ 850 5 ~ 873 4 ~ 910 911 ~ 918	47 48 49 40
36	CWL	RELEASING	WALKING	920 ~ 931	55	M_M	RELEASING	B _WALKING	919 ~ 934	51
37	CWL	CWL	WALKING	932 ~ 957	56	M_M	M_M	F_WALKING	935 ~ 957	52

Quantitative Evaluation of Autonomous Segmentation Framework [5/5]

episode [ID0-0]

episode [ID0-2]

episode [ID0-11]

episode [ID0-12]

	ID0-0	ID0-2	ID0-11	ID0-12
# of dimension reduced by PCA	2	1	3	1
# of Basis Skills autonomously segmented by our method	68	52	83	47
# of Segments manually segmented by A. Yao	99 (38)	56 (13)	97 (23)	55 (10)
# of similar segments	60 (7)	37 (5)	58 (18)	37 (7)
Similarity of Segments	88.24%	71.15%	69.88%	78.72%
Similarity of Segments	98.53%	90.38%	91.57%	93.62%

# of segments with 1~5frames

# of segments which have different granularities

when *timing differences between* ***the starting and ending points*** *in the segments are allowed* ***from 0 to 10frames*** (i.e. 0.0~0.4 sec)

when ***identically considering the segments that have the difference of segmentation granularities*** and ***eliminating the segments with 1~5frames*** (it is difficult to find physical meaning)

Dissimilar primitives can be easily explained by the difference of segmentation granularity that can be considered in motions such as opening, closing, and walking.

INCORL **In**telligence and **Con**trol for **R**obots **L**aboratory

Reminding:
How can we reuse primitives well?

- Reorganization of New Sentences using Words

sentences "HEISABOY" "SHEISNOTAGIRL"

words HE IS A BOY | SHE IS NOT A GIRL

reorganization

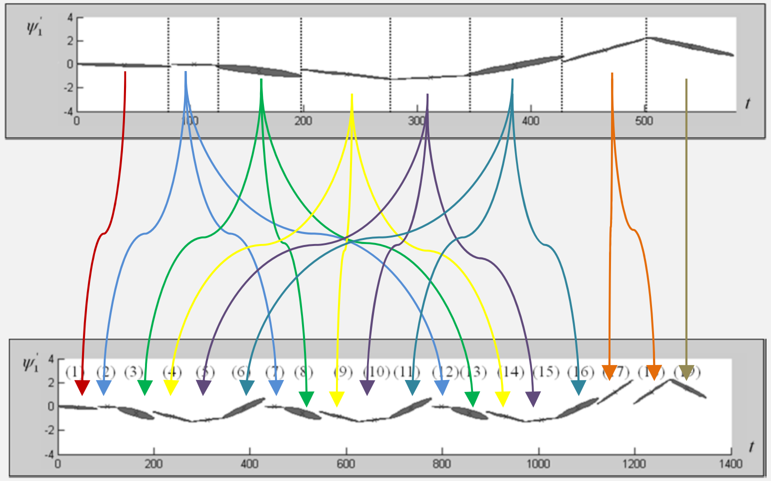
New sentence "HE IS NOT A GIRL"

Reusability → Primitives (Words) | Pre- and Post-conditions

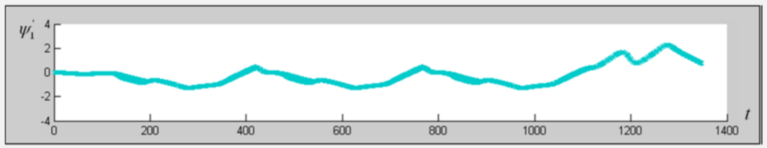
Reorganization of Primitives Learned from a Single Task

Reuse of Primitives Learned by Imitation

< GMM that consists of eight Gaussians
and temporally overlapping points
in-between consecutive Gaussians >



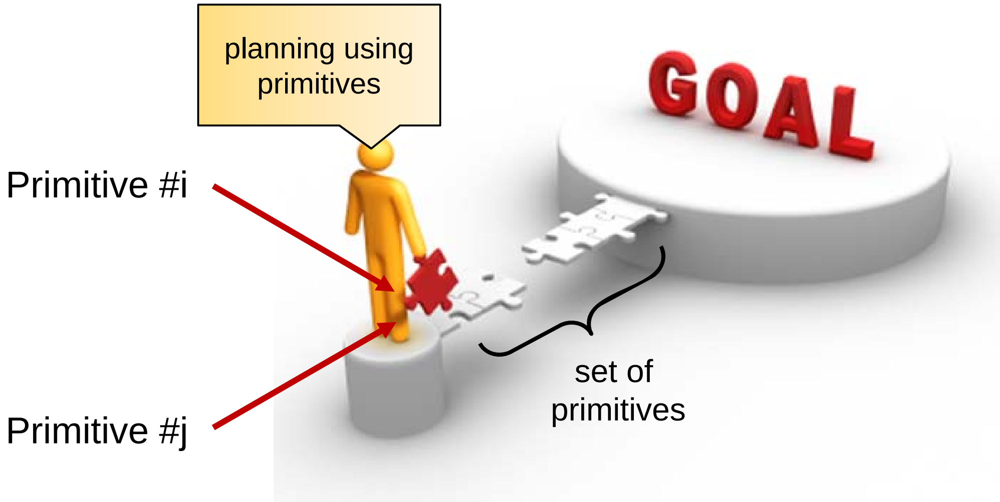
[Gaussian Mixture Model in the dimensional space reduced by PCA]



[Gaussian Mixture Model in the dimensional space reduced by PCA]

[the task of three scooping, three delivering,
and two stirring rice]

Grammaticalization of Primitives

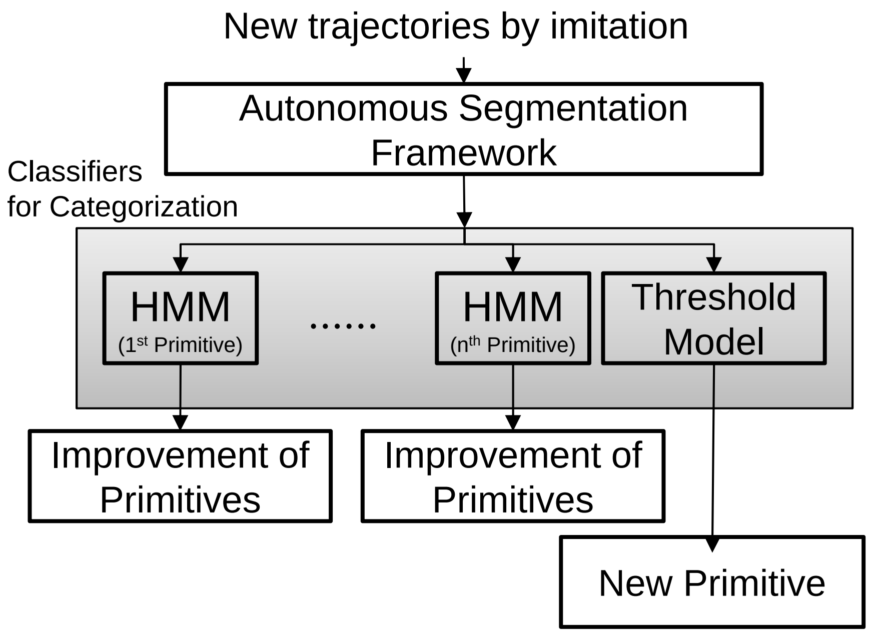


Crucial Requirements
Categorization and Generalization

Grammaticalization of Primitives

Simple Approach for Categorization

- Categorization **: Hidden Markov Model**

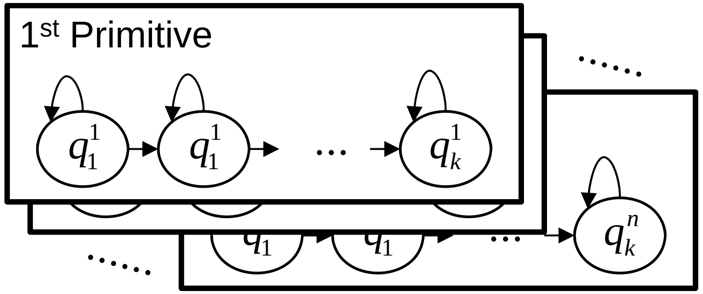


Grammaticalization of Primitives

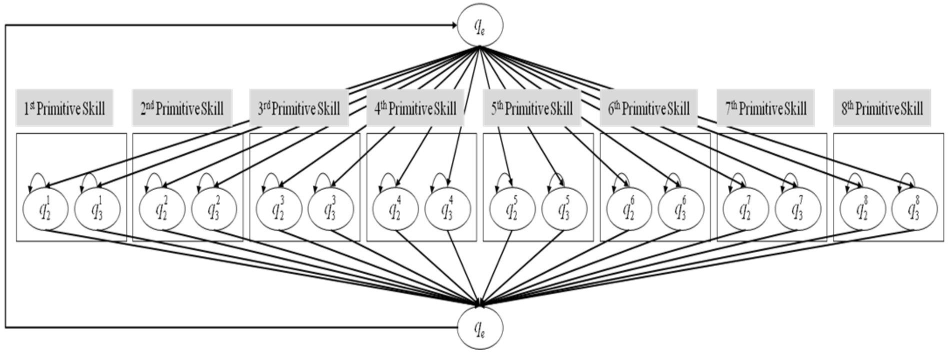
Simple Approach for Categorization

- Categorization **: Hidden Markov Model**

< HMM Representation of Primitives >



< Threshold Model >



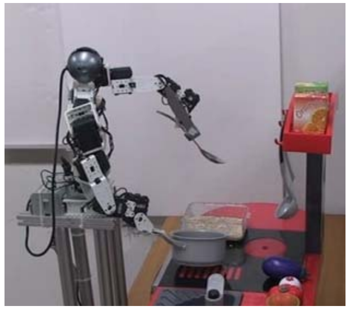
ergodic HMM using HMM states of existing HMMs

Grammaticalization of Primitives

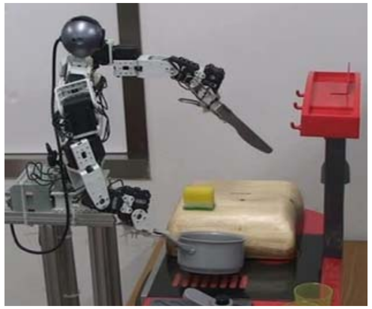
Simple Approach for Categorization

- Categorization **: Hidden Markov Model**

< The Same Category >



[LiftingSpoon]



[LiftingKnife]

Eights Segments : [LiftingPot], [LiftingSpoon],
(cooking rice) [ApproachingRiceBowl], [ScoopingRice],
[DeliveringRice], [PouringRice],
[StirringRice], and [PuttingOnStove].

< Threshold Model >

[CuttingFoodItem]

[PositioningForPushing]

[PushingFoodItem]

Reorganization of Primitives Learned from Multiple Tasks

Reuse of Primitives Learned from two tasks of cooking rice and cutting a food item

original sequence in the cooking rice

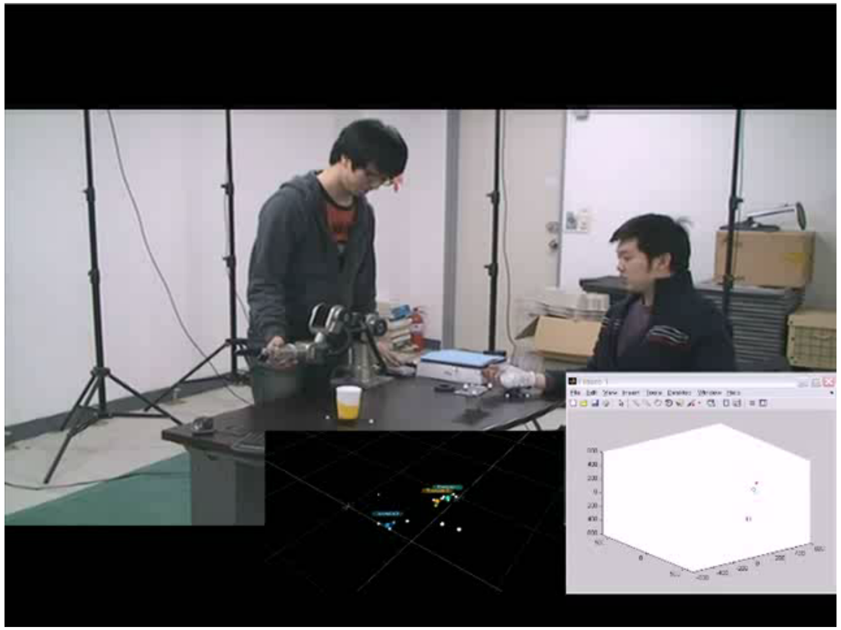
original sequence in the cutting a food item

[the task of two cutting, one pushing, two stirring, and one putting on the stove]

Can the segmentation points be used for gramaticalization?

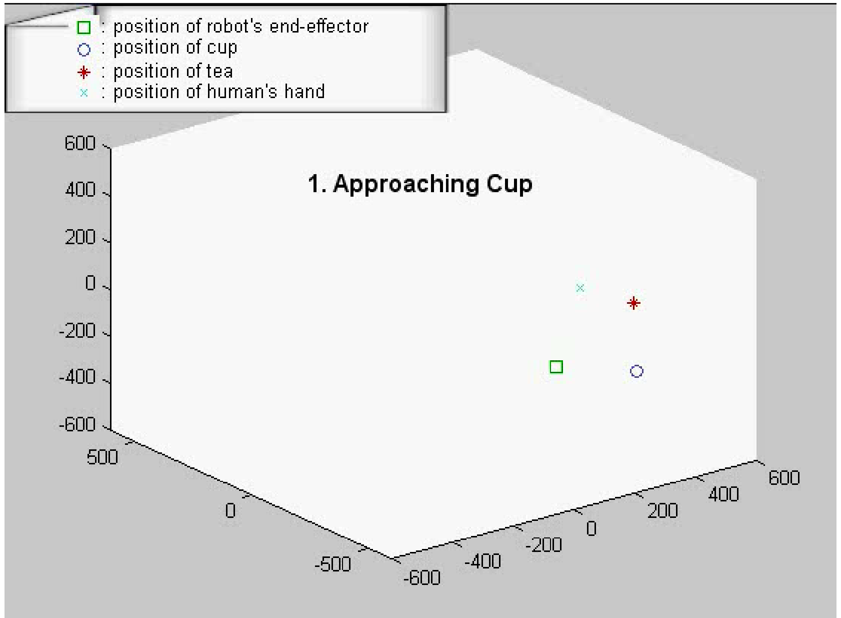
- When segmenting spatial information of surrounding objects using the segmentation points

< A demonstration of preparing tea >



[00:00:26]

< Segmentation results by autonomous segmentation framework >



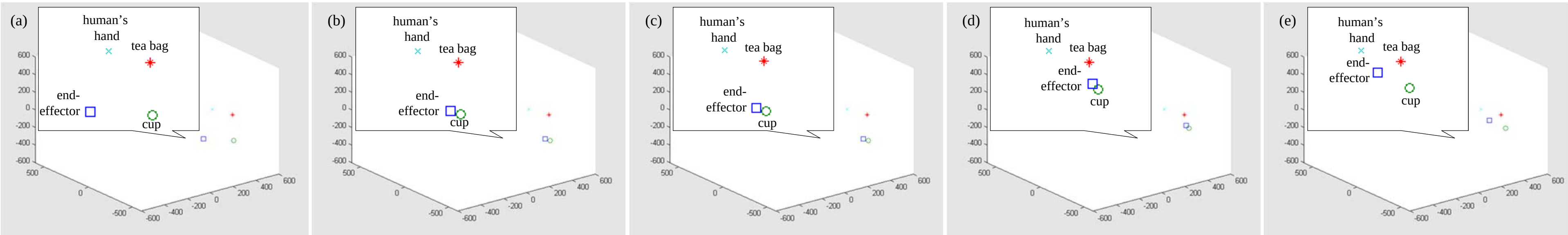
[00:00:48]

6-D robot arm developed by Neuronics (i.e. Katana)
12 motion capture cameras developed by Optitrack (i.e. V100:R2)

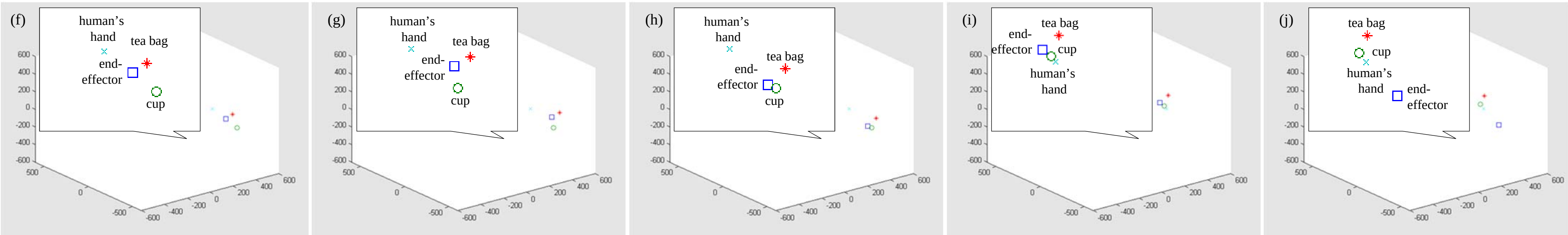
.aboratory

Can the segmentation points be used for gramaticalization?

- Illustrations captured in the nine segmentation points



[InitialPositions] [ApproachingCup] [GraspingCup] [DeliveringCup] [ApproachingTeaBag]



[GraspingTeaBag] [InsertingTeaBag] [GraspingCup] [DeliveringCup] [ComingBack]

The segmentation points can be sufficiently used to determine pre- and post-conditions to activate primitives

Future Works

- Rich Representation for Proto-language to Categorize and Generalize Primitives
  - Affordances
  - Object Action Complexes
  - Motion Algebra
- Key question remaining
  - “Whom to imitate”, “When to imitate”, and “What to imitate”
  - How can we evaluate the learning performance?

Thank you !!!

Challenge
: Grammaticalization of Primitives [5/6]

Grammaticalization (including Categorization)

- In linguistics,
  - a process by which words representing objects and actions (i.e. nouns and verbs) transform to become grammatical objects (e.g., affixes and prepositions etc.)
- In Robotics (especially, behavior),
  - a process in which information representing objects and actions (i.e. conditions and behaviors (or primitives)) transforms (categorizes and relates) to become grammatical objects for planning

Hidden Markov Model

: Efficient Method to Categorize Primitives

Affordances or Object Action Complexes (OACs)

: Method to Categorize and Grammaticalize Primitives, simultaneously

[Papers]

[1] N. Kruger, C. Geib, J. Piater, R. Petrick, M. Steedman, F. Worgotter, A. Ude, T. Asfour, D. Kraft, D. Omercen, A. Agostini, and R. Dillmann,, "Object-Action Complexes: Grounded abstractions of sensory-motor processes," RAS, 59(10), pp.740-757, 2011.

[2] F. Worgotter, A. Agostini, N. Kruger, N. Shylo, B. Porr, "Cognitive agents-a procedural perspective relying on the predictability of Object-Action-Complexes (OACs)," Robotics and Autonomous Systems, 2008.

[3] E. Sahin, M. Cakmak, M. Dogar, E. Ugur, and G. Ucoluk, "To afford or not to afford: A new formalization of affordances toward affordance -based robot control," Adaptive Behavior, pp.447-472, 2007.

[4] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, "Learning object affordances: From sensory-motor coordination to imitation," IEEE Trans. on Robotics, 2008.

…

[Projects]

1. PACO-Plus Project (2006 ~ 2010) : FP 6
2. Xperience (2011 ~ 2015) : FP 7

Challenge
: Grammaticalization of Primitives [5/6]

Classical Plan Operator Representation

(behavior, (pre-conditions, effect))

e.g., STRIPS operators

ex1) (index : swim
action : swim
precondition: river,
effect: traversed)

ex2) (index : walk
action : walk
precondition: road,
effect: traversed)

Affordance Representation

(effect, (entity, behavior))

e.g., Affordance relations

ex1) (index : traversed
effect : traversed
(entity: river, behavior : swim)
(entity: road, behavior : walk)

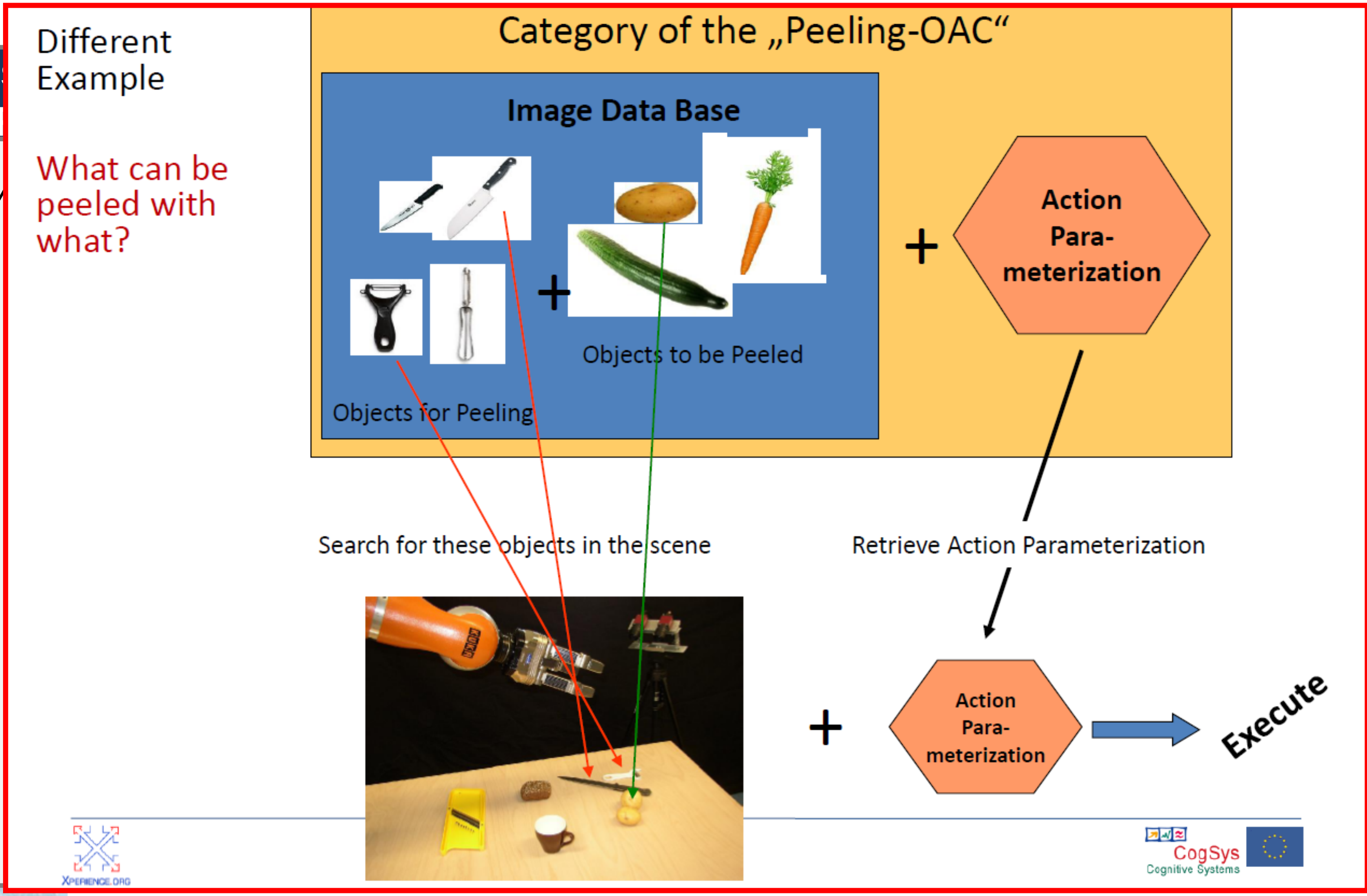
OACs Representation

$$execute\ \underbrace{(E, T, M)}_{\textbf{OAC}} \rightarrow (s_0, s_p, s_r)$$

E: an identifier for an execution specification
T: a prediction function of how the world will change after executing *E*
M: a statistical measure representing the success of the OACs
s_0 : the state of the world before performing OAC
s_p : the state of the world that T predicted from OAC
s_r : the observed state from actually performing *E*

ex1) Name : ObjGrasp
Attribute space/*T* : Object model, gripper status
M : long term probability of successful grasp

Appendix – Example of OACs



Clas
I
(behav
e.g.,
ex1)
ex2)

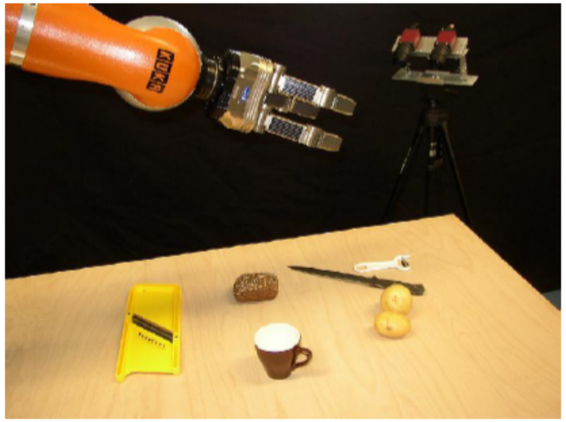
tation
s_0, s_p, s_r)
pecification
e world will
ting
performing
redicted from
lly performing
ctor's pose
bject location
be
ediction
n

Appendix – Example of OACs

For Example asking the robot:

What can be cut with what?

(without having seen any of the objects before!)

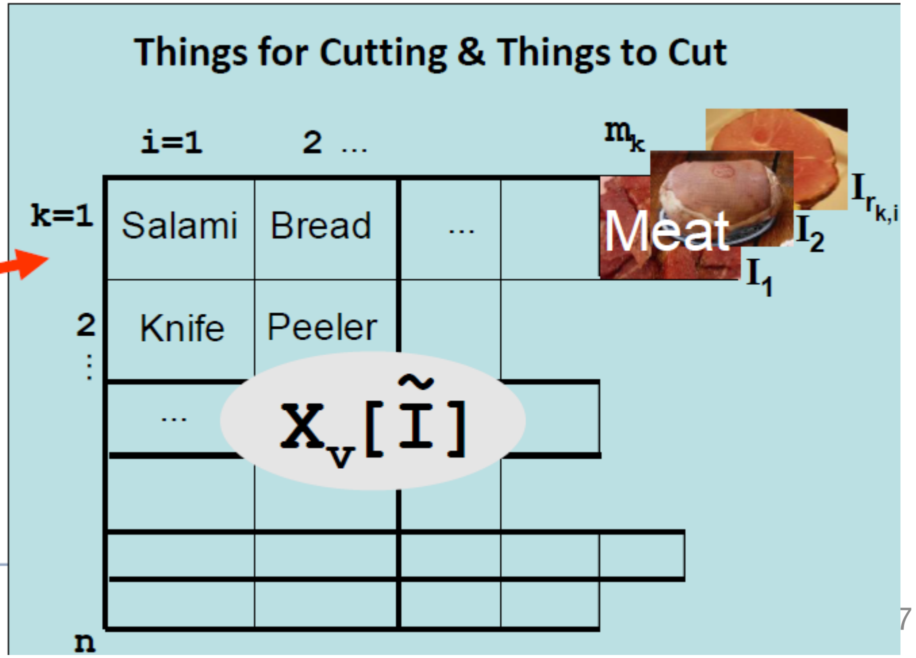


Algorithm: Generalize, starting with the sentence:

"**Cut the salami with a knife**"

use the Internet to **replace nouns** in this sentence and then **attach images** to the new nouns (again from the internet) .

Store a verb-labeled "**Picture Book**" of what can be cut with what.



Challenge
: Grammaticalization of Primitives [5/6]

Classical Plan Operator Representation

(behavior, (pre-conditions, effect))

e.g., STRIPS operators

ex1) (index : swim
action : swim
precondition: river,
effect: traversed)
ex2) (index : walk
action : walk
precondition: road,
effect: traversed)

Affordance Representation

(effect, (entity, behavior))

?

?

e.g., Affordance relations

ex1) (index : traversed
effect : traversed
(entity: river, behavior : swim)
(entity: road, behavior : walk)

OACs Representation

execute (E, T, M) → *(s_0, s_p, s_r)*

?

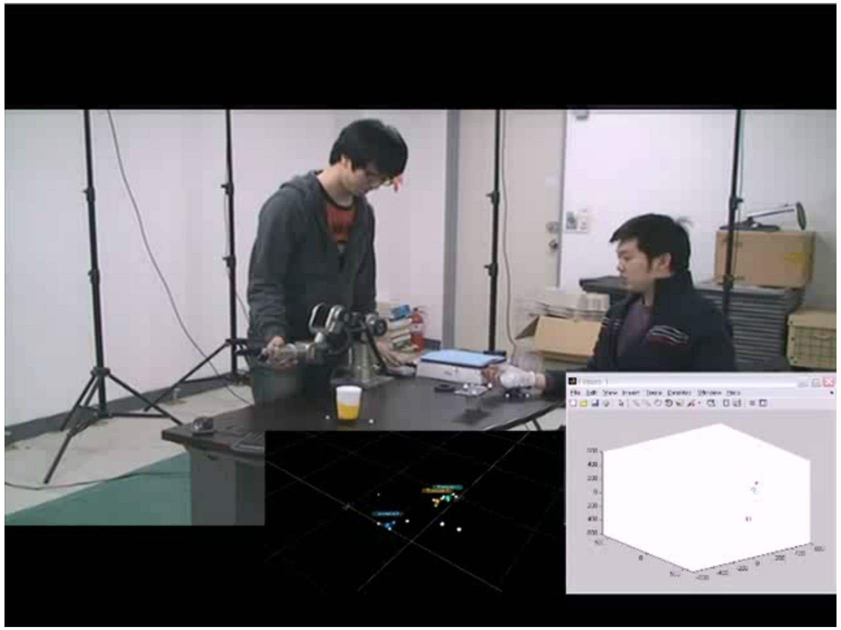
E: an identifier for an execution specification
T: a prediction function of how the world will change after executing *E*
M: a statistical measure representing the success of the OACs
s_0 : the state of the world before performing OAC
s_p : the state of the world that T predicted from OAC
s_r : the observed state from actually performing *E*

ex1) Name : AgnoPush
Attribute space/*T* : end effector's pose space, object location and shape
M : average deviation of prediction from actual final position

Challenge
: Grammaticalization of Primitives [6/6]

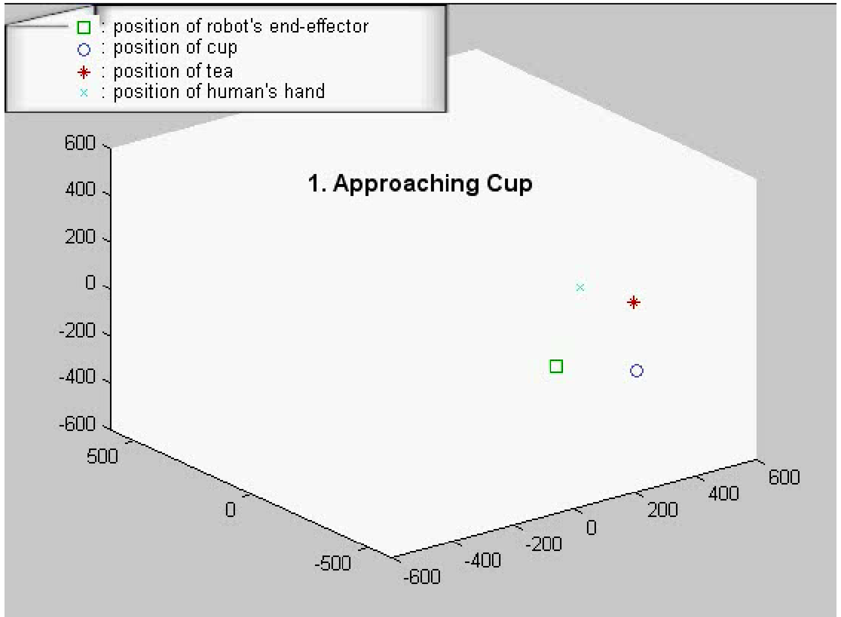
- The Potential Possibility in which the Segmentation Points can be used to Determine Pre- and Post-conditions

< A demonstration of preparing tea >



[00:00:26]

< Segmentation results by autonomous segmentation framework >

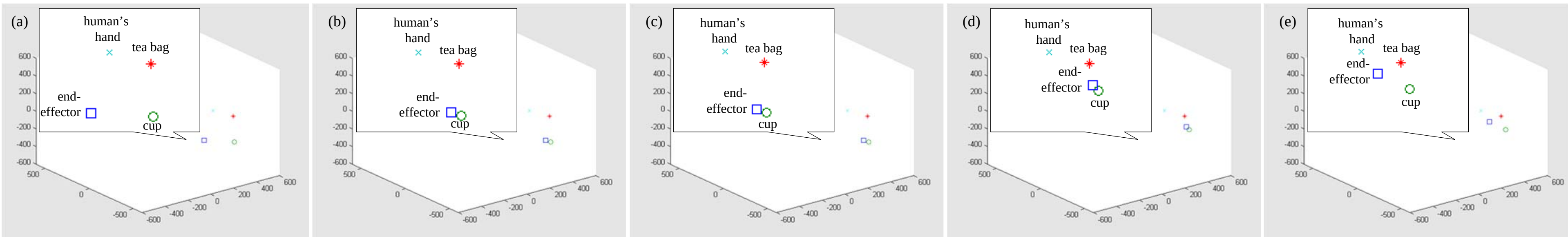


[00:00:48]

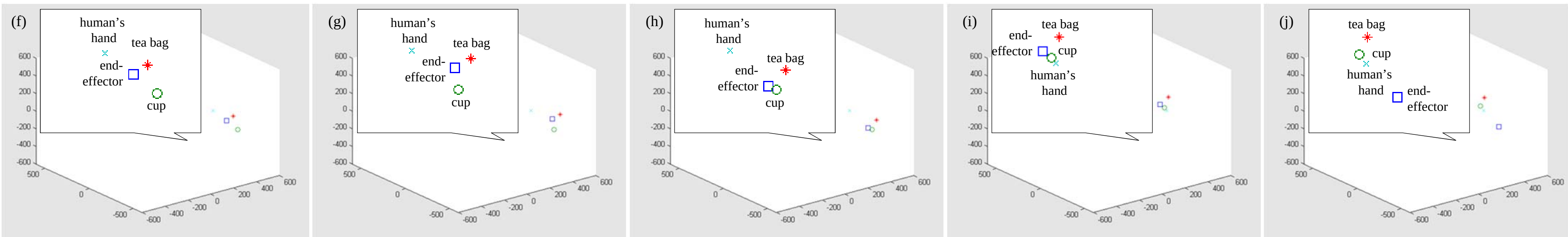
Challenge

: Grammaticalization of Primitives [6/6]

- The Potential Possibility in which the Segmentation Points can be used to Determine Pre- and Post-conditions



[InitialPositions] [ApproachingCup] [GraspingCup] [DeliveringCup] [ApproachingTeaBag]



[GraspingTeaBag] [InsertingTeaBag] [GraspingCup] [DeliveringCup] [ComingBack]

Challenge
: Grammaticalization of Primitives [5/6]

Classical Plan Operator Representation

(behavior, (pre-conditions, effect))

e.g., STRIPS operators

ex1) (index : swim
action : swim
precondition: river,
effect: traversed)

ex2) (index : walk
action : walk
precondition: road,
effect: traversed)

Affordance Representation

(effect, (entity, behavior))

e.g., Affordance relations

ex1) (index : traversed
effect : traversed
(entity: river, behavior : swim)
(entity: road, behavior : walk)

Our Primitives

Using surrounding information in segmentation points

Extension of Affordance and OACs

: (effect, (entity, behavior), M)

OACs Representation

execute (E, T, M) → (s_0, s_p, s_r)

...fier for an execution specification
...ion function of how the world will
...after executing *E*
...cal measure representing
...ess of the OACs

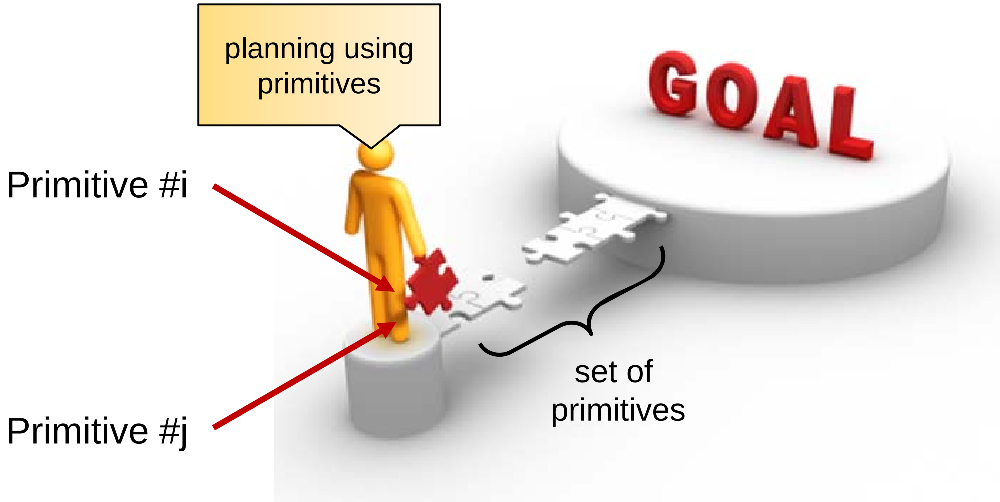
s_0 : the state of the world before performing OAC

s_p : the state of the world that T predicted from OAC

s_r : the observed state from actually performing *E*

ex1) Name : AgnoPush
Attribute space/*T* : end effector's pose space, object location and shape
M : average deviation of prediction from actual final position

Challenge
: Grammaticalization of Primitives [1/6]



Crucial Requirements

For Categorization and Generalization

: Grammaticalization

Examples In “Xperience” Project – FP 7

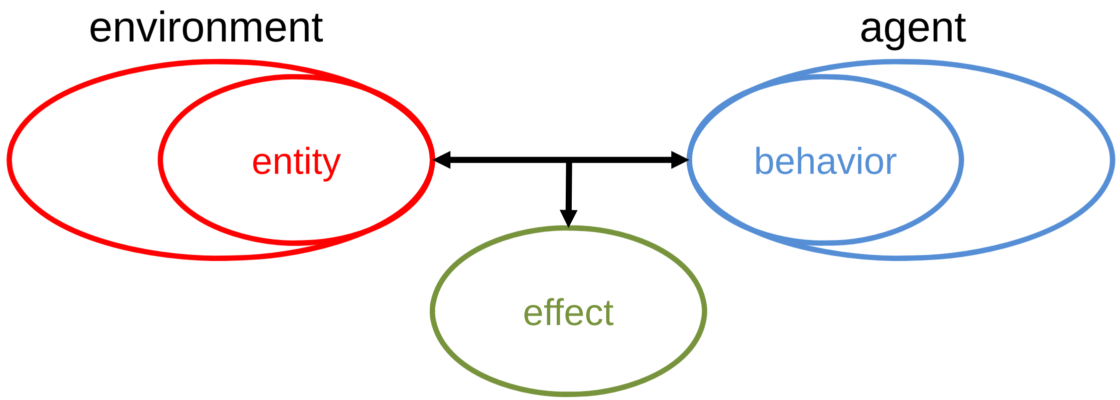
- In Language domain,
  - knowing the **grammar** of English and the **category** and **meaning of the surrounding words in a sentence** allows identification of **the category and semantic type of an unknown word**.

- In Robotics (sensorimotor) domain,
  - knowing how to **peel** potatoes with a **knife**, significantly aids one in learning how to use a potato-peeler. A single demonstration enables understanding in terms of an existing theory of **potato peeling**, and makes **the peeler** available for **generalization to other plans** (other potatoes and other vegetables).

APPENDIX I

Challenge
: Grammaticalization of Primitives [6/13]

affordance



(effect, (entity, behavior))

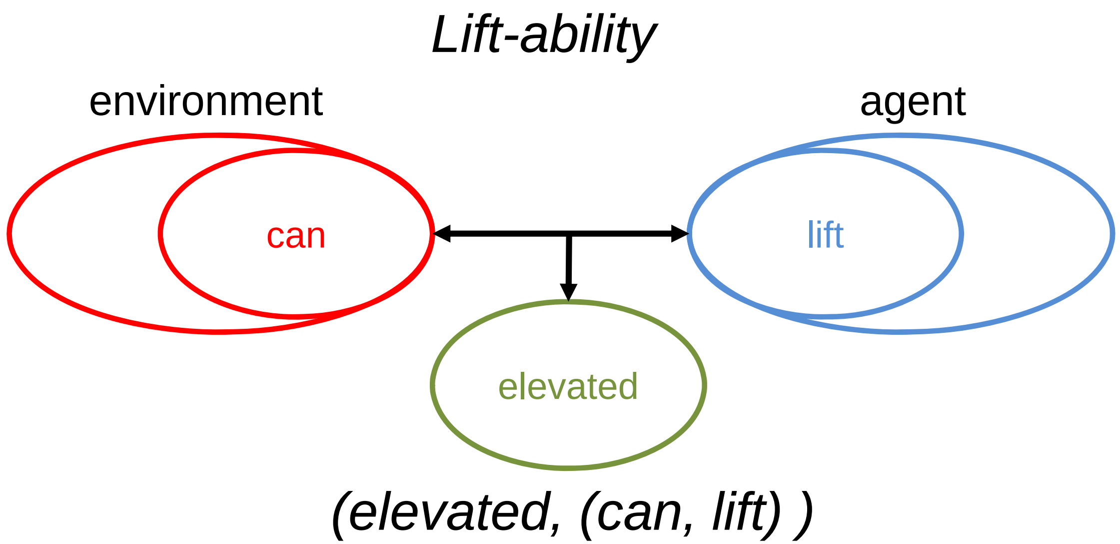
Definition of "Affordance"

- An acquired ***relation between a behavior*** (i.e. a primitive skill) of an agent and ***an entity*** in the environment such that the application of the behavior on the entity generates ***a certain effect***.

< E. Sahin, M. Cakmak, M. R. Dogar and E. Ugur, "To Afford or Not to Afford: A new formalization of Affordances toward Affordance-based Robot Control," Adaptive Behavior, December, 2007>

Challenge
: Grammaticalization of Primitives [7/13]

- Example of Affordance in Robotics



"Lift-ability"

- The robot applied its lift behavior on the can and obtained the elevated effect.

  Can: the perceptual representation of the can as seen by the robot

  Lift : the behavior executed by the robot

  Elevated : the effect of the behavior on the environment as perceived by the robot

< E. Sahin, M. Cakmak, M. R. Dogar and E. Ugur, "To Afford or Not to Afford: A new formalization of Affordances toward Affordance-based Robot Control," Adaptive Behavior, December, 2007>

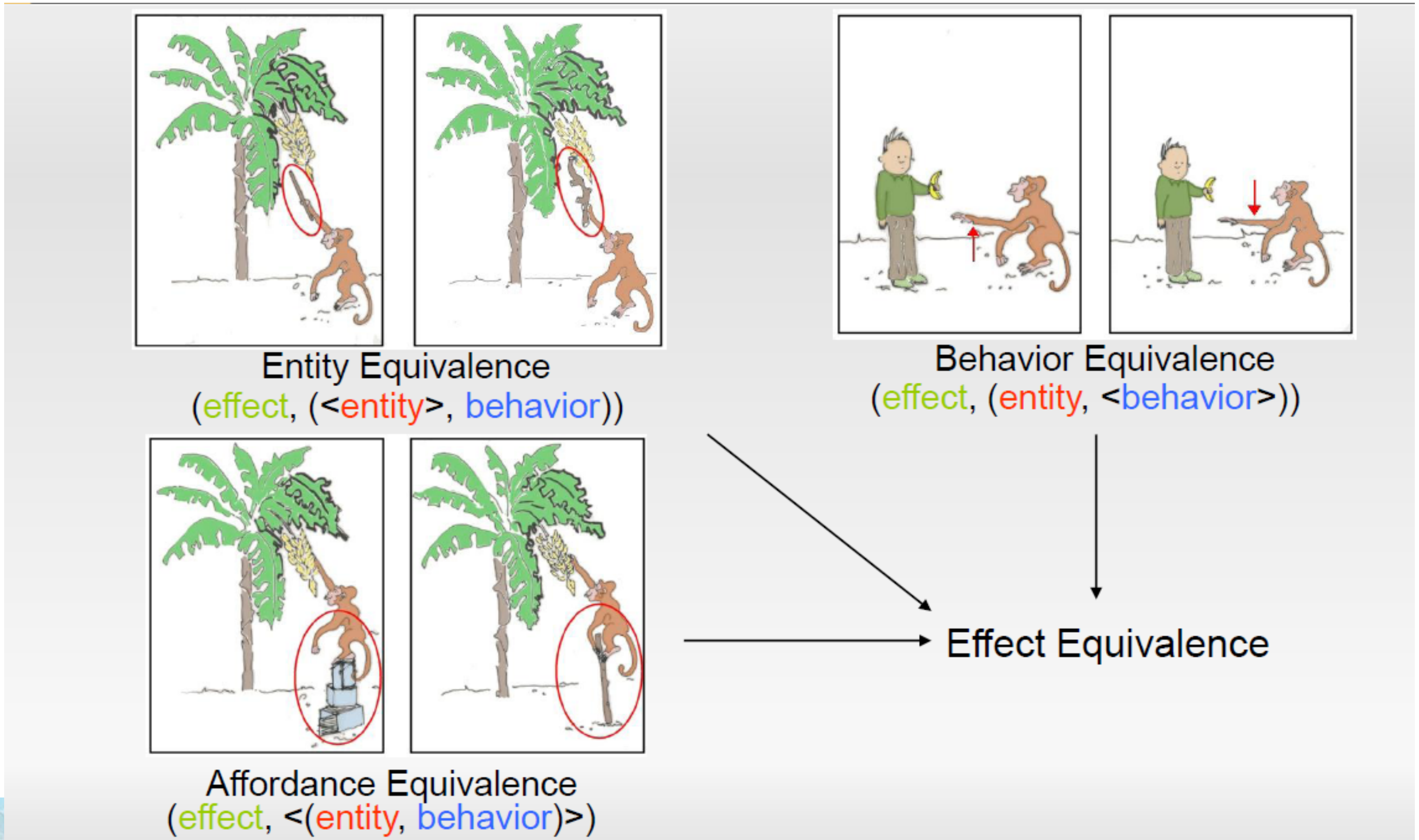
Challenge
: Grammaticalization of Primitives [8/13]

Classical Plan Operator Representation	Affordance Representation
(behavior, (pre-conditions, effect))	*(effect, (entity, behavior))*
e.g., STRIPS operators	e.g., Affordance relations
ex1) (index : swim action : swim precondition: river, effect: traversed) ex2) (index : walk action : walk precondition: road, effect: traversed)	ex1) (index : traversed effect : traversed (entity: river, behavior : swim) (entity: road, behavior : walk)

< E. Sahin, M. Cakmak, M. R. Dogar and E. Ugur, "To Afford or Not to Afford: A new formalization of Affordances toward Affordance-based Robot Control," Adaptive Behavior, December, 2007>

Challenge
: Grammaticalization of Primitives [9/13]

- Strategy of Categorization : **Effect Equivalence**

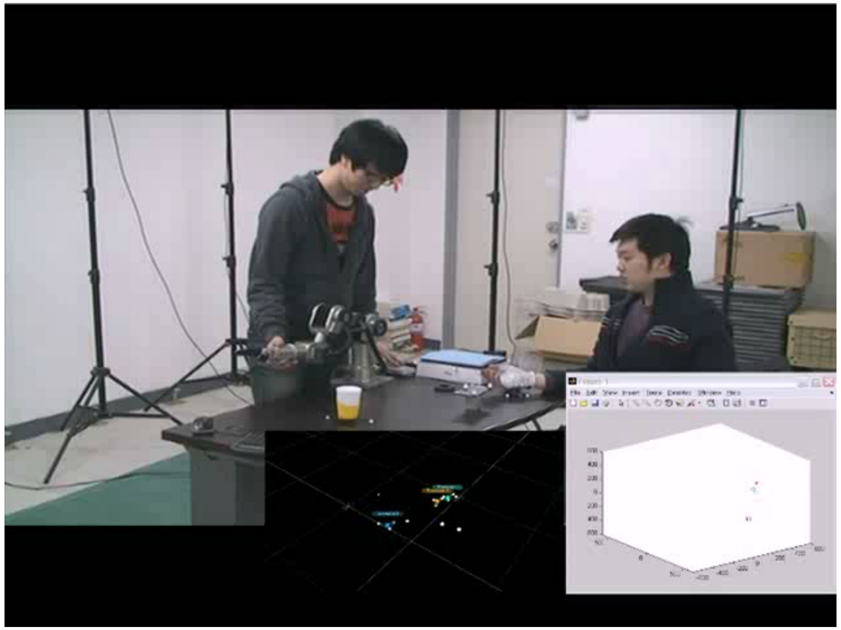


< E. Sahin, M. Cakmak, M. R. Dogar and E. Ugur, "To Afford or Not to Afford: A new formalization of Affordances toward Affordance-based Robot Control," Adaptive Behavior, December, 2007>

Challenge
: Grammaticalization of Primitives [10/13]

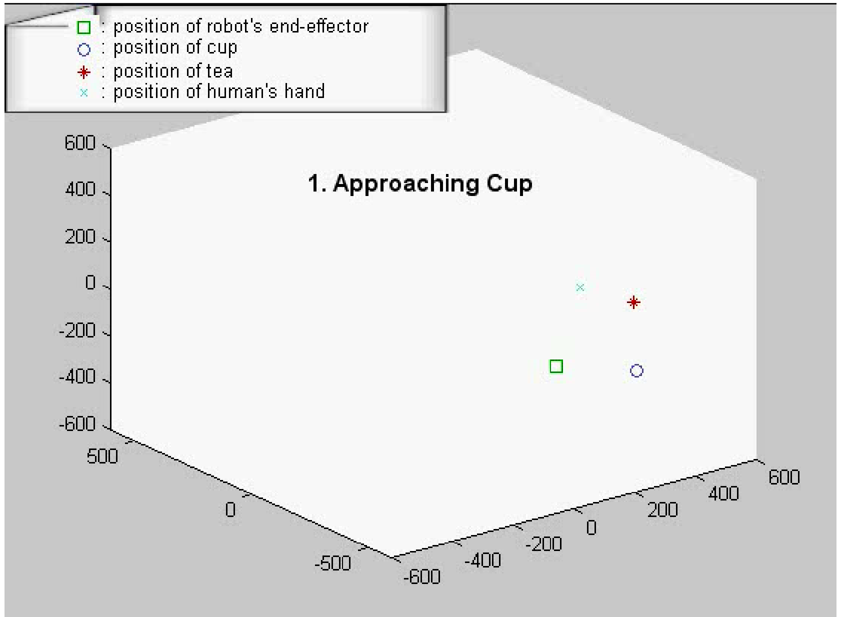
- Are there affordances or effect equivalence in the task of preparing Tea?

< A demonstration of preparing tea >



[00:00:26]

< Segmentation results
by autonomous segmentation framework >



[00:00:48]

Challenge
: Grammaticalization of Primitives [11/13]

- In Task of Preparing Tea : a naïve example

Effect Equivalence using Symbolization of Effects

Using Feat

1. relative distan
2. relative distan
< x-axis, y-axis,

pre

1. < -296.35 ,

2. < -15.58, 6

Symbolizat

1. < -296.35 ,
entity: < -1 ,

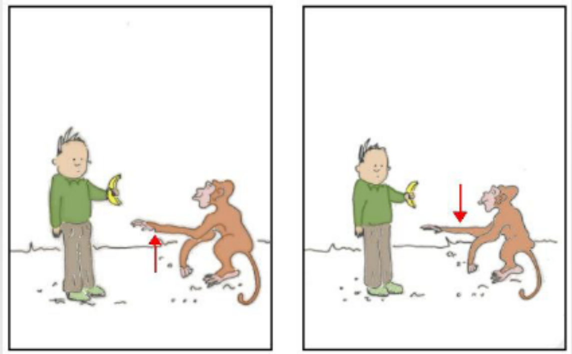
2. < -15.58, 6
entity: < -1, +

affordance # *i*

(index: pattern < +1, -1, -1 >
effect: pattern <+1, -1 -1 >
entity: pattern < -1, +1, +1 >, behavior: primitive #1)
entity: pattern <-1, +1, +1 >, behavior: primitive #4)

Categorization
by Effect Equivalence

Grammaticalization
for Planning



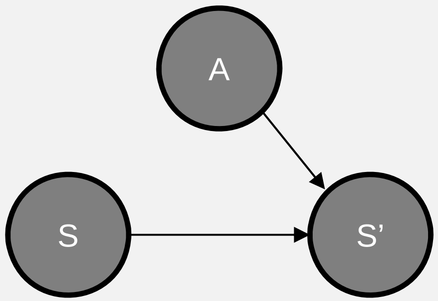
Behavior Equivalence
(effect, (entity, <behavior>))

Challenge
: Grammaticalization of Primitives [12/13]

Extending Affordance Representation for a Robot

affordance # i

(index: pattern < +1, -1, -1 >
effect: pattern <+1, -1 -1 >
entity: pattern < -1, +1, +1 >, behavior: primitive #1)
entity: pattern <-1, +1, +1 >, behavior: primitive #4)



[Bayesian Network]
: e.g., Naïve Bayes Classifier

Probabilistic Representation using Real Values

(entity: <+1, -1, -1 > , (entity: < -1, +1, +1 >, behavior: primitive #1))
→ ***< -296.35 , 45.99, 143.2016 > , primitive #1 , < -2.26, 30.99, 1.75 >***

(entity: <+1, -1, -1 > , (entity: < -1, +1, +1 >, behavior: primitive #4))
→ ***< -15.58, 69.52, 81.08 >, primitive #4, < -13.13, 55.07, 23.96 >***

Challenge
: Grammaticalization of Primitives [13/13]

Original Affordance Representation

(effect, (entity, behavior))

(index: pattern < +1, -1, -1 >
effect: pattern <+1, -1 -1 >
entity: pattern < -1, +1, +1 >, behavior: primitive #1)
entity: pattern <-1, +1, +1 >, behavior: primitive #4)

Extended Affordance Representation

(effect, (entity, behavior), BN)

(index: pattern < +1, -1, -1 >
effect: pattern <+1, -1 -1 >
entity: pattern < -1, +1, +1 >, behavior: primitive #1, prob._model: BN #1)
entity: pattern <-1, +1, +1 >, behavior: primitive #4, prob._model: BN #4)

APPENDIX II

Object Action Complexes (OACs)

- **OACs** are proposed as a **universal representation enabling efficient planning** and **execution of purposeful action** at all levels of a cognitive architecture.
- **OACs** combine the r**epresentational and computational efficiency for purposes of search** (the frame problem**) of STRIPS rules** and **the object- and situation-oriented concept of affordance** with the **logical clarity of the event calculus**.
- While **affordances** have mostly been analyzed in their purely **perceptual aspect**, the **OACs** concept defines them **more generally as state transition functions suited to prediction**.
- Such functions can be used for **efficient forward planning, learning, and execution of actions** represented simultaneously at multiple levels in an embodied agent architecture.
  - PACO+ proejct , FP6 (2006~2010), Xperience project ,FP7 (2011~2015)
  - **Objects and Actions are inseparably intertwined** in cognitive processing; that is “Object-Action Complexes” (OACs) are the building blocks of cognition.
  - **Cognition is based on reflective learning, contextualizing and then reinterpreting OACs to learn more abstract OACs**, through a grounded sensing and action cycle.
  - The core measure of **effectiveness** for all learned cognitive structures is: Do they **increase** situation **reproducibility** and/or **reduce** situational **uncertainty** in ways that allow the agent to achieve its goals?

[1] Krüger, N., Piater, J., Wörgötter,F., Geib, Ch., Petrick, R., Steedman, M.; Ude, A., Asfour, T., Kraft, D., Omrcen, D., Hommel, B., Agostino, A., Kragic, D., Eklundh, J., Kruger, V. and Dillmann, R.(2009). A Formal Definition of Object Action Complexes and Examples at different Levels of the Process Hierarchy.

[2] Wörgötter, F., Agostini, A., Krüger, N., Shylo, N. and Porr, B. Cognitive agents - a procedural perspective relying on the predictability of Object-Action-Complexes (OACs). Robotics and Autonomous Systems, 2008.

[3] Geib, Ch., Mourao, K., Petrick, R., Pugeault, N., Steedman, M., Krüger, N. and Wörgötter, F. Object Action Complexes as an Interface for Planning and Robot Control. IEEE-RAS International Conference on Humanoid Robots (Humanoids 2006).

[4] Justus Piater, Mark Steedman, Florentin Wörgötter. Learning in PACO-PLUS.

[5] Retrieved from "http://en.wikipedia.org/w/index.php?title=Object_Action_Complex&oldid=478584468"

XPERIENCE

Robots Bootstrapped through Learning from Experience

Rüdiger Dillmann

XPERIENCE.ORG
Jan 2011 - Dec 2015

Karlsruhe Institute of Technology Germany
R. Dillmann, T. Asfour

University of Göttingen, Germany
F. Wörgötter

University of Innsbruck, Austria
J. Piater

Italian Institute of Technology, Italy
G. Metta, G. Sandini

University of Southern Denmark
N. Krüger

Jozef Stefan Institute, Slovenia
A. Ude

University of Edinburgh, United Kingdom
M. Steedman, C. Geib

Xperience: Problem and Approach

- **State of the Art (developmental approach):** Exploration of the world allows acquiring grounded and robust cognitive representations. This is an "outside-in", data-driven process.
- **Human cognitive ability:** We are able to also use generative mechanisms based on (e)Xperience for knowledge extension.
  - **Advantage:** This is an "inside-out", model-driven process and much faster!

Approach: XPERIENCE will implement a complete robot system combining developmental with generative mechanisms for automating introspective, predictive, and interactive understanding of actions and dynamic situations.

Structural Bootstrapping

- The process of structural bootstrapping compares a newly observed entity to a model of experienced entities to understand the novel situation and predict consequences of actions.

- The concept is taken from human language acquisition
  - Example: Knowledge of "Fill a bottle with water", allows you to infer the role of xxx as something that can be filled with water when hearing the sentence "Fill the xxx with water".

- Xperience transfers this concept to the full spectrum of cognitive robotics problems.

Examples for Structural Bootstrapping

1. Language domain: Knowing the grammar of English and the category and meaning of the surrounding words in a sentence allows identification of the category and semantic type of an unknown word.

2. Sensorimotor domain: Knowing how to peel potatoes with a knife, significantly aids one in learning how to use a potato-peeler. A single demonstration enables understanding in terms of an existing theory of potato peeling, and makes the peeler available for generalization to other plans (other potatoes and other vegetables).

Major Scientific Questions

1. How to improve exploration based knowledge acquisition ("outside-in" stage)?
2. How to implement the generative process of structural bootstrapping ("inside-out" stage)?
3. How to combine these two mechanisms in a dynamically stable process?
4. How to predict other agents, leading to advanced abilities to cooperate, interact and communicate?
5. How to integrate a complete embodied cognitive system?

OACs as representations in Xperience

- Object-Action Complex (OACs, pronounced "oaks")
  - Grounded abstractions of sensorimotor processes
  - Describes how an object is affected by an action
  - Can be executed to actually do it
  - Allows reasoning based on experience
  - Combines notions of
    - affordances (perception)
    - prediction (action, state transitions)
    - reasoning (~STRIPS)
- OACs as basis for symbolic representations of sensorimotor experience and behavior.

Krüger et al. 2011. Object–Action Complexes: Grounded abstractions of sensory–motor processes, RAS, 59(10):740-757, 2011

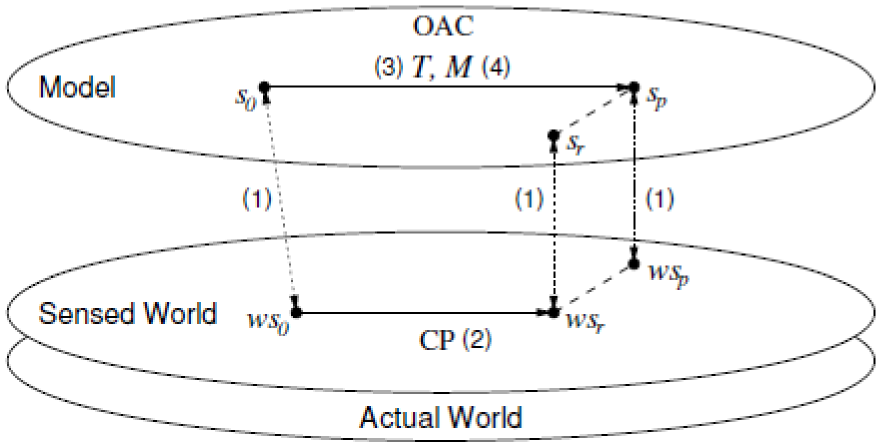


Figure 3: Graphical representation of the OAC learning problems: (1) Translation, (2) Control, (3) Prediction, and (4) Reliability.

$$(E, T, M) \tag{1}$$

where:

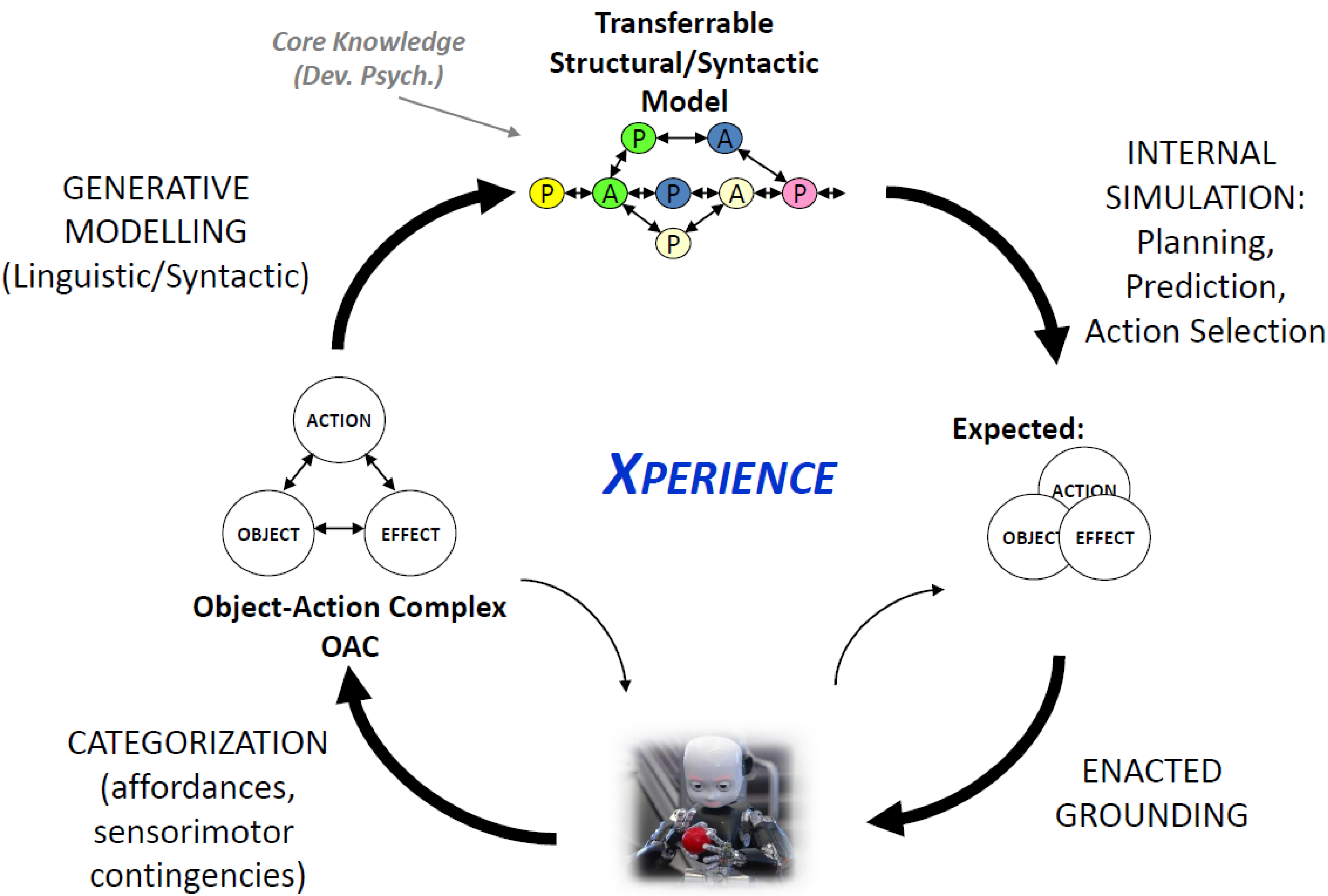
- *E is an identifier for an execution specification,*
- *$T : \mathcal{S} \rightarrow \mathcal{S}$ is a prediction function defined on an attribute space $\mathcal{S}$ encoding a model of how the world (and the agent) will change if the execution specification is executed, and*
- *M is a statistical measure representing the success of the OAC in a window over the past.*

$$\texttt{execute} : (E, T, M) \rightarrow (s_0, s_p, s_r), \tag{2}$$

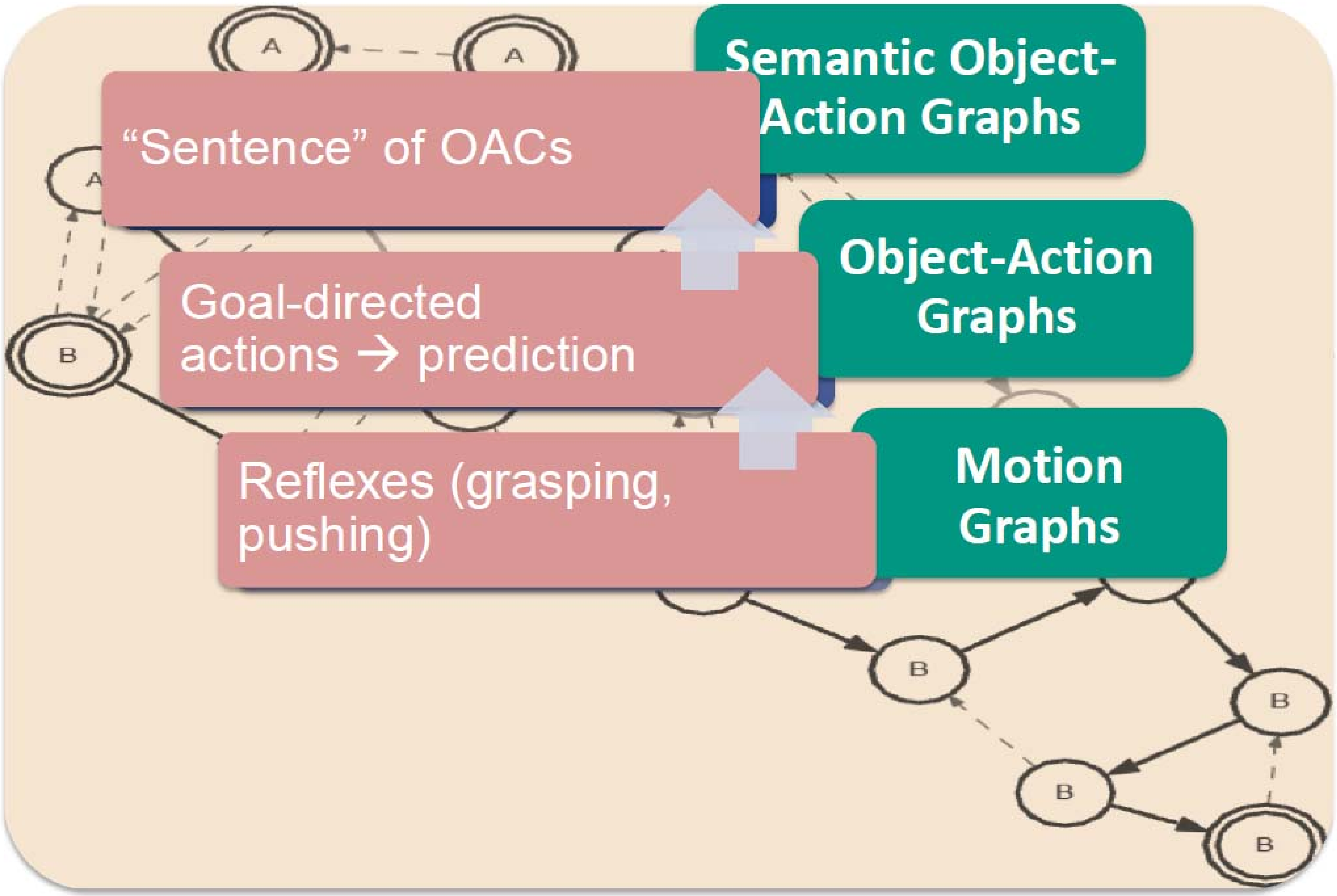
where:

- *$s_0 \in \mathcal{S}$ is the state of the world before performing the OAC's execution specification,*
- *$s_p \in \mathcal{S}$ is the state of the world that T predicts will result from performing the OAC's execution specification in s_0, i.e., $s_p = T(s_0)$, and*
- *$s_r \in \mathcal{S}$ is the observed state resulting from actually performing E in state s_0.*

The XPERIENCE Cognitive Architecture

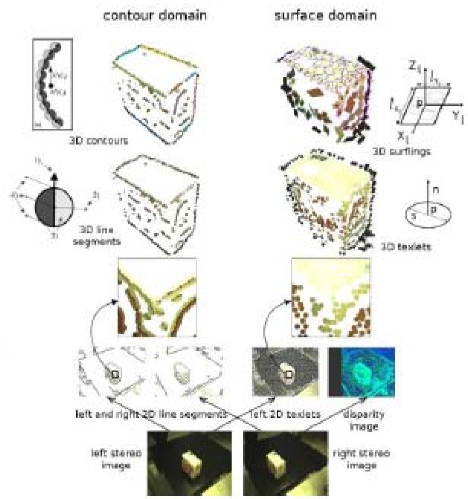


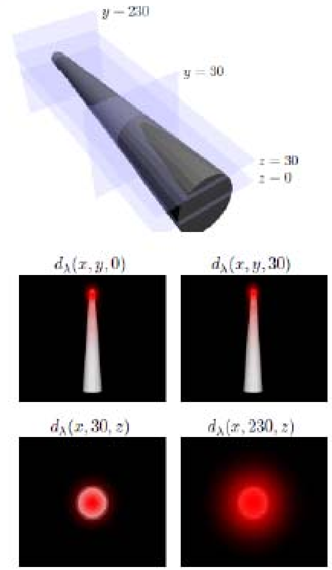
OACs on all levels



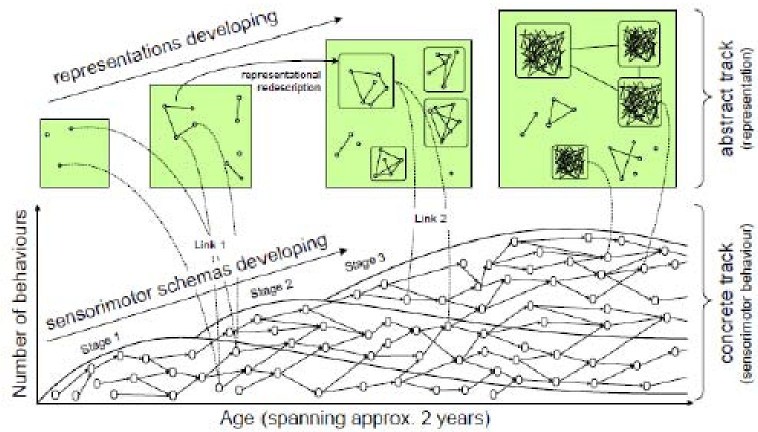
Learning hierarchical and probabilistic sensory-motor spaces: Early Cognitive Vision (ECV) x Probabilistic Grasp Functions (PMFs)

- ECV provides
  - a deep hierachical, view point invariant, rich, explicit visual representation
- PMFs
  - provide a probabilisitc, complete and structured action representation
- OACs
  - provide the required framework for generating, storing and utilizing sensory-motor data
- Structural booststrapping on a sensory-motor level
  - searches in the cross space ECV x MD for relevant structures
  - to refine existing and create new OACs





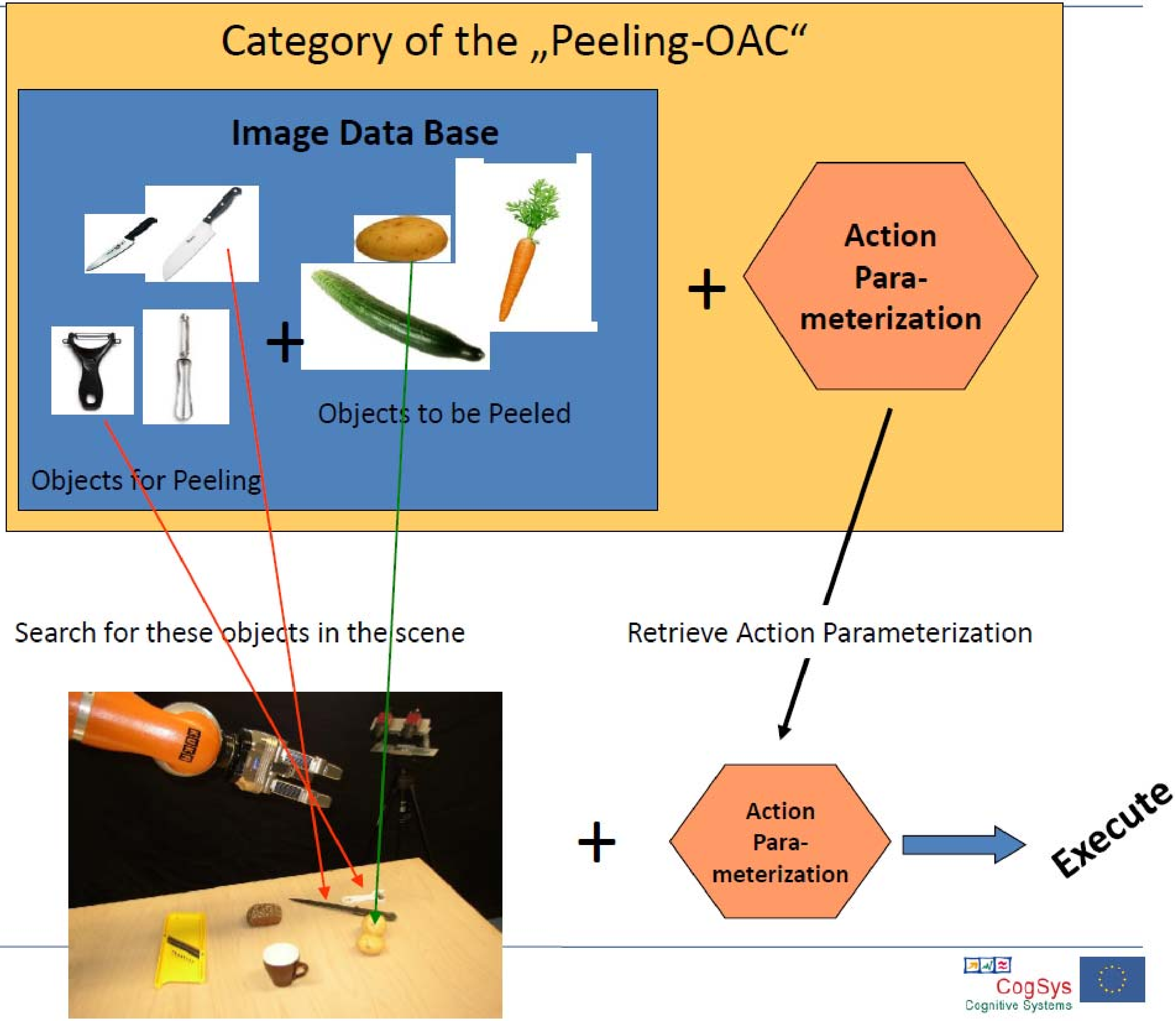
ECV **x** **MD**



Generalizing Objects by Analyzing Language (“GOAL”)

Different Example

What can be peeled with what?

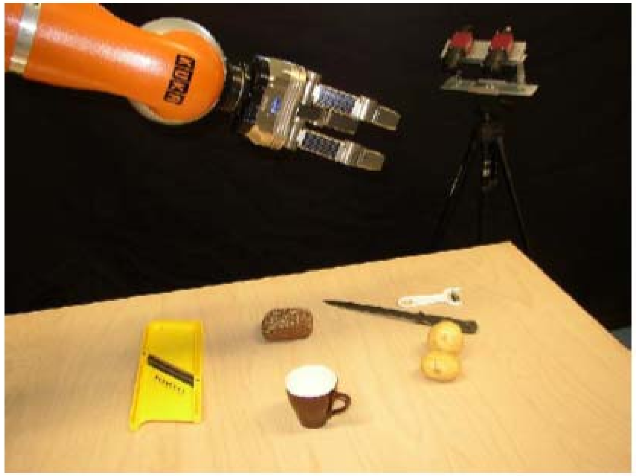


Generalizing Objects by Analyzing Language ("GOAL")

For Example asking the robot:

What can be cut with what?

(without having seen any of the objects before!)



<u>Algorithm:</u> Generalize, starting with the sentence:

"**Cut the salami with a knife**"

use the Internet to **replace nouns** in this sentence and then **attach images** to the new nouns (again from the internet) .

Store a verb-labeled "**Picture Book**" of what can be cut with what.